Technical University of Munich

Master's Thesis

The search for predictability and correlations between traffic-data based congestion- and incident characteristics – An exploratory data analysis

Die Suche nach Vorhersagbarkeit und Korrelationen zwischen verkehrsdatenbasierten Stau- und Ereignismerkmalen – Eine explorative Datenanalyse

Jakob Erpf

Mentoring:

M. Eng. Barbara Karl (TUM)

Dr. - Ing. Matthias Spangler (TUM)

Dipl. - Ing. Stefan Gürtler (S&W)

Dipl. - Ing. Johannes Grötsch (LBD)

December 2, 2020

Department of Civil, Geo and Environmental Engineering

Kurzfassung

Aktuelle Navigationssysteme verwenden häufig Akkumulationsstrategien, um die Reisezeit abzuschätzen, wobei Zeitverzögerungen durch Staus auf der Grundlage der Analyse der Geschichte des zugrunde liegenden Straßennetzes berücksichtigt werden. Dieser Ansatz kann durch ungewöhnliche Ereignisse gestört werden, die kurze Zeitblockaden verursachen, oder durch regelmäßig anfallende, langfristige Verringerungen des Verkehrsaufkommens verzerrt werden. Diese Arbeit evaluiert einen neuen Ansatz zur Vorhersage von Stau- und Unfallmerkmalen durch die Korrelation von Staus und Ereignissen, die in zeitlicher und räumlicher Nähe zueinander liegen. Um dies in einer explorativen Datenanalyse zu evaluieren, werden drei reale Datensätze aus dem Jahr 2019 betrachtet, die Verkehrsbewegungs- und Störfalldaten liefern. Nach einem algorithmischen Ansatz zur Detektion von Staus in FCD Daten und der Lokalisierung räumlich und zeitlich benachbarter Vorfälle aus den BAYSIS und ArbIS Datensätzen wird in der Arbeit mit statistischen Methoden evaluiert, ob und wie diese Vorfälle und Staus miteinander korreliert sind. Daher besteht die Methodik aus dem Clustering von FCD, dem Matching benachbarter Vorfälle, der Korrelations- und Vorhersageanalyse.

Die Ergebnisse zeigen, dass es signifikante Korrelationen zwischen Stau- und Störfallmerkmalen gibt, was bedeutet, dass einzelne Unfallmerkmale statistisch zu Staus mit einer bestimmten Länge und Dauer führen. Dieser Zusammenhang zwischen Länge und Dauer eines Staus ist auch bei den Baustellencharakteristika, wie der Baustellenlage (Straße) und des Baustellenausführungsmonat gegeben. Obwohl diese Korrelationen einen ersten Hinweis auf die Vorhersagbarkeit liefern, ergab eine separate Analyse, dass zwischen keinem der Merkmale Vorhersagbarkeit besteht. An dieser Stelle muss auch angemerkt werden, dass viele Beziehungen nach der Klassifizierung der Daten keine ausreichende Stichprobengröße mehr hatten und ein größerer Datensatz notwendig wäre, um mehr aussagekräftigere und zuverlässigere Ergebnisse zu finden.

Abstract

Current navigation systems often use accumulation strategies to estimate travel time while considering time delays through congestions based on analyzing the history of the underlying street network. This approach can be disturbed through uncommon events creating short time blockages or be biased through regular accruing, long-term traffic volume reductions. This thesis evaluates a new approach of predicting congestion and accident characteristics through the correlation of congestions and incidents which are placed in temporal and spatial vicinity of each other. To evaluate this in a exploratory data analysis, three real world datasets from the year 2019 will be considered providing traffic movement and incident data. After an algorithmic approach to analyzing a derivative of a floating car dataset for jams and locating spatial and timely adjacent incidents from the BAYSIS and ArbIS, the thesis will evaluate if and how these incidents and jams are correlated with each other with statistical methods. Therefore the methodology consist of the clustering of FCD, adjacent incident matching, correlation and prediction analysis.

The results show that there are significant correlations between congestion and incident characteristics which means that individual accident characteristics statistical lead to jams with a certain length and duration. This length and duration relationship is also present with the roadwork characteristics of the roadwork's location (road) and the month of the roadwork. Although these correlations provide a first indication of predictability, a separate analysis revealed not predictability between any of the characteristics. At this point it also has to be noted that many relations had insufficient sample size after classifying the data and a larger dataset would be necessary to find more significant and reliable results.

Erklärung zur Master's Thesis

"Ich versichere hiermit, die vorliegende Arbeit selbständig verfasst und keine anderen Quellen als die angegebenen Quellen und Hilfsmittel benutzt zu haben. Die Arbeit wurde noch nicht anderweitig für Prüfungszwecke vorgelegt."

München, 15.12.2020 : ______________________________

B. Sc. Jakob Erpf

Contents

1 Introduction

Traffic jams are a common problem to everyone, ever attempting to begin the summer season with a car trip on the first day of summer break. When the number of road users increases or the capacity of the road way decreases due to various reasons, a imbalance between demand and supply is created (Tang et al. 2019). This impacts passenger traffic as well as transportation of goods through blocked bottlenecks and decreased travel speeds. They lead to unreliable travel times, inefficient usage of resources, and an increase in emissions, like pollutants or noise (FHA 2011). Another effect is a decrease in road safety, due to high driver tempers or inattentive mind, which can result in higher accident counts (Sun et al. 2016). This induces enormous social costs due to billions of hours lost in jams and induced mental stress. (ADAC 2019; Bardt et al. 2014; Retallack et al. 2019)

Therefore it is essential to reduce risks of jams, as well as accidents, with an increased understanding of the traffic accident causes, trigger effects of jams and roadwork consequences, to maintain a fluent traffic flow and traffic safety. Tailored Transportation System Management (TSM) strategies, focused on automatic reactions for significant traffic event, could enable Active Traffic and Demand Management (ATDM) of high traffic demands in reduced traffic volume areas (Tang et al. 2019). This would go towards reducing the economic, environmental, and social costs associated with accidents, roadworks or jams. Part of these TSM strategies implemented in a ATDM system, could be a form of traffic incident prediction systems, with the potential to identify compromising conditions in real-time, allowing according to actions to be taken to avoid consecutive events. (Retallack et al. 2019)

This thesis approaches congestion and incident prediction by evaluating the statistical relation of jams and incidents to predict the chance of a consecutive event. These consecutive events can be jams, as well as incidents. Depending on the severity of an accident jams can be provoked. On the other hand jams can facilitate accidents due to the change of traffic flow. Another scenario are construction sites and maintenance which can also lead to both jams and accidents, because of the reduction of traffic volume, changes in road guidance, or other modifications to the actual traffic situation. Although the scope of the thesis does not cover the specifics on a complete production system for congestion and accident detection, prediction and response, it will take the concept of such a system and focus the possibility of predicting such events, which then would make the development of such a system possible.

A system capable of the described mentioned functionalities would likely consist of the following processing components.

- Acquisition of traffic movement and incident data
- Congestion and incident detection
- Prediction of consecutive events
- Traffic management and controlling responses

The second component in charge of detecting jams as well as incident, requires input data like speed, volume, occupancy to represent the traffic situation and incident reports to define incident characteristics. The next component would then analyze the resulting dataset to find characteristic features of the jams and incidents, which will be determined in this thesis. In case the analysis of the characteristic shows a possible and imminent event it would trigger the last component to initiate appropriate controlling actions and prepare incident responses. In the following chapters of this introduction section, the reader will be introduced to the concepts and systems used in this these to cover the input and output requirements of such ATDM system components.

1.1 Continuous Floating Data

To detect jams continuous data about the speed or movement of the vehicle on the road is necessary. This kind of information can be collected through a variety of different systems to represent a real-time or at least current picture of the traffic situation.

The current street network of Bavaria heavily depends on stationary sensors to assess the traffic situation. This includes inductive loops, infrared or radar detectors which can provide traffic indicators like volume, speed, time gaps, jams, density and many others. The data collected with just stationary sensors can only describe the traffic trends restricted to their location and coverage which requires complex simulations and modeling to aggregate data for the missing areas where no sufficient coverage can be provided. Adding to this is the fluctuating result set quality which depends on the input data and simulation model quality. Especially highways are equipped with stationary sensors but the lower index streets network is only covered by a fragmented net of detectors with distances of up to 100 km between detectors (INDRIX 2015). Fortunately, nowadays cars as well as drivers and riders are equipped with technology that allows real time tracking and comprehensive data collection. Automobiles can gather information from the build in sensors as well as the on-board GPS. Even mobile devices from drivers and riders can be used to collect location and movement data. (Randelhoff 2016)

Floating Car Data (FCD) is continuously collected during the usage of a car by the on-board GPS and represents the individual movements. Typically this incorporates the coordinates, timestamp, road section, course and routing data points. These are regularly sent to a central FCD unit/service via mobile radio communication (GSM-, UMTS- or LTE-based) to be aggregated and combined with stationary data to an area wide picture of the traffic situation.

In this form they can be used for traffic analysis and management. (LAPID 2020; Randelhoff 2016)

A derivative of FCD is Extended Floating Car Data (XFCD) developed by BMW. It expands the collection of data points of an FCD with data from the vehicle sensors and systems like breaks, rain sensors, driver assistance systems and more. These data points add a number of analytic opportunities to generate a more precise and detailed traffic picture. (LAPID 2020)

In contrast to FCD, Floating Phone Data does not need an on-board GPS or car systems to create movement data but assumes that driver's and rider's mobile devices will register and deregister at the radio tower along the road. FPD uses this registration information to determine the radio cell the phone is currently in, how long is stayed in that cell and tracks the devices when switching to another cell or tower. It is therefore able to collect a much larger quantity of datasets but lacks in the accuracy of FCD which transmit the GPS location of the car itself. That been said, on roads with a high coverage of radio tower like in urban areas or on highways, FPD is able to achieve a comparable accuracy. Mobile service providers collect this anonymized FPD and forward them to a traffic management unit, which can analyze the date for disturbances and give feedback through the traffic information channels. (LAPID 2020; Randelhoff 2016)

Another type of floating data type is Floating Car Observer (FCO). FCO does not only collect it's own FCD but also data about it's surroundings with the built-in sensors. This includes the automatic recognition of cross- or opposite traffic, traffic volume or relative speeds to other cars. This additional data not only add detail but also allow for correctional fusion algorithms to reduce uncertainties or errors in the pure FCD. (Randelhoff 2016)

1.2 Street Information Systems

Besides of having a real-time or current picture of the traffic situation, the concept of the thesis also relies on information about current disturbances or possible triggers of disturbances.

Bayerische Straßeninformationssystem (BAYSIS)

For disturbances in form of accidents the Bayerische Straßeninformationssystem (BAYSIS), a publicly available information system from the Bavarian street administration, will be used. The systems task is the acquisition, collection and analysis of street network related information, which contains infrastructure inventory and condition, traffic volumes and other key values, as well as an accident register with detailed reports. An export of this accident register with detailed reports is provided by the Zentralstelle Verkehrsmanagement (ZVM) for this thesis.

Arbeitstellenintegrationsystem (ArbIS)

Another type of disturbance to consider is roadworks, for which an export from the Arbeitstellenintegrationsystem (ArbIS), a software tool and database used by the Bavarian infrastructure ministry, will be consulted. The system is used to collect and archive all current, planned and passed roadwork and maintenance projects on the Bavarian state street net-work. With the collected information from ArbIS, an effective, economic and safe execution of roadworks can be achieved. Furthermore ArbIS can provide detailed report exports about current projects to the Bavarian traffic information office and traffic information channels. (Trafficon 2017)

1.3 Traffic Status Information

To deliver information from traffic management systems to the road users traffic messaging channels can be used. Three known examples are Real Time Traffic Information (RTTI), Traffic Messaging Channel (TMC) and Advanced Traveler Information System (ATIS).

The TMC is a messenger system for jams and other traffic incidents. The public available service is free to use and publishes current congestion notifications in compatible navigation system and announcements in the local radio channels. In the scope of the TMC the road network is split into TMC sections. These sections are used to define the spatial location of the incident to report and typical start or end with road linkups. More detailed Incident information is obtained from the police or reports from traffic participants which adds a considerable before publication delay. It is therefore spatial and temporal to rather imprecise. (LAPID 2020)

Another traffic information source is the RTTI. RTTI supplies traffic participants with information about current events or suggested diversions like Traffic Messaging Channel (TMC), but with a much spatial higher accuracy. Through a *Geocast*, which is an expansion of a multicast with a geolocation, the spatial precision of the RTTI is superior to the TMC. This *Geocast* can be either a geometrical address like a GSM84 coordinate or a symbolic address, reaching a spatial accuracy of up to 100m. (Hinden et al. 2006; Imielinski et al. 1996; LAPID 2020). This is why RTTI is the industry standard of suppling vehicle and third party navigation system with up to date traffic information. Another difference to TMC is the accessibility. Unlike the publicly available TMC, RTTI is vendor specific and most of the time a payed service, like BMWs ConnectedDrive (BMW 2020).

The most up-to-date variation of a traffic status information system is the Advanced Traveler Information System (ATIS), such as GoogleMaps, HERE or Waze. Like RTTI it is a provider-based service, but mostly without costs and less device constraint. Because of these missing accessibility constraints and the fact that in current times nearly 70% of people carry a smartphone (IZM 2020), the user base of such services is quite substantial. This does not invalidate the usage of RTTI and TMC, since they are present in most separate and built-in navigational systems, but rather makes ATIS a considerable alternative. With the added benefit of being able to not only publish information about the traffic situation but also collect variations of FCD, ATIS is the most promising technology for ATDM and FCD collection.

In the scope of this thesis only the statistical relations of the input congestion and incident data is evaluated. Therefore the traffic response channel is not relevant further, but the following system:

- FCD for the congestion data
- BAYSIS and ArbIS for incident data

The thesis is based on the analysis of three datasets, which are associated a confidentiality agreement for the usage in this thesis and therefore can not be added to the appendix. To provide an complete picture of the data analysis and also additional information to interested reader mosts scripts, results and visual representations are available in a version controlled open source repository[1]. The submitted versions of the thesis also contain the confidential code which can not be made publicly available.

[1]https://github.com/jakoberpf/master-thesis

2 Objects of Research

This chapter defines the objects or events which will be researched in this thesis. This includes a literature review and descriptive definition of a congestion. For the use in the data analysis jam events will be further defined based on FCD.

2.1 Incident

Incidents in the scope of this thesis can be accidents, as well as ongoing roadwork or maintenance on the Bavarian street network. These are also the events, which the concept of the thesis tries to predict through the analysis of the correlation of said incidents to jams.

Accident

An accident is an unexpected and unintentional traffic event, that typically results in damages, injuries or reduction of traffic volumes. These events can be triggered by a number of different reasons, where in this thesis the main focus is on the triggers of slow, congested traffic or roadworks.

Roadwork

All static and moving construction sites classify as roadwork, as well as temporary blockages or disturbance due to snow clearing, road maintenance and alike.

2.2 Congestion

Naming As there does not exist a clear plural of the noun congestion, the term jams will be used in case of multiple congestion. These two terms are seen as interchangeable for their reference to a single or multiple congestion events.

2.2.1 Descriptive Definition

Jams or a single congestion in layman terms are spatial and temporal accumulations of traffic participants, resulting in speeds slower or sometimes much slower than free flow. In severe cases this is also often described as stop-and-go or stopped traffic. They are triggered by a reduction of traffic throughput in volume or an increase of traffic demand. Studies have shown that these triggers are usually caused by four categories of disturbances, which are defined by the *Transportation Research Board* (TRB 2003) and *Federal Highway Administration* (FHA 2011).

Traffic-Influencing Events

- Incidents : Events that disrupt the normal free flow of traffic, like vehicular crashes, breakdowns or debris. These physical obstacles block lanes or hard shoulders, forcing other road users to execute evasive maneuver and deviate from their normal path. This ultimately changes driving behavior, reduces the quality of traffic flow and traveling speed. Even when incidents are not directly on the roadway they can impact the traffic flow due to emergency responses that create blockades or ineffective driving behavior of traffic participants gaping on the incident.
- Roadwork : Managed and unmanaged construction sites on the roadway that result in physical changes to the highway environment. This includes a reduction of lanes, lane diversion, elimination of hard shoulders or road closures, which reduce the road capacity and reduce travel speeds.
- Weather : Changes in environmental conditions like heavy rain or snow fall can negatively impact driver behavior. The reduction of visibility will usually result in a reduction of traveling speeds and increase of headway. This reduces the overall capacity of the highway. Bright sunlight, smoke or icy road surfaces lead to a similar effect.

Traffic Demand

- Fluctuations in Normal Traffic : Variations in demand in day-to-day traffic volumes can overload systems with fixed capacities. This can result in travel speed reductions without any specifically occurring events.
- Special Events : Special cases where events drastically change the demand in their vicinity and overload the system. As with incidents, off-road events can affect driving behavior due to visual distractions and change the traffic-flow.

Physical Highway Features

- Traffic Control Devices : Poorly timed or defective traffic signals as well as other ineffective traffic flow control contributes to the creation of jams and travel time increase.
- Physical Bottlenecks or Capacity : The capacity of a road is mostly dependent on the number of lanes and hard shoulder, as well as the alignment (curves and grades). Physical changes on the road environment like in merging areas, tool booths or road endings reduce the capacity and therefore promote the formation of jams. The road capacity can also be influenced by the driving behavior, which heavily depends on the familiarity of the roadway to the driver. Drivers familiar with routinely congested road tend to reduce their headway and therefore increase the capacity (Charlton et al. 2013).

Driving-behavior

As mentioned above, driving behavior can influence the traffic flow as well as capacity and is mostly influenced by the environment and the familiarity of the road. Research showed that driving on familiar roads has a negative effect on safety aspects of driving behaviors, like unintentional blindness for roadside features (ibid.). Another decreasing factor is the state-of-mind, better known as rage- or aggressive driving, resulting in rapid lane changing, cross cutting or passing on shoulders (Shinar et al. 2004). This can lead to driving behaviors where drivers do not keep up smooth accelerations, but rather break suddenly or accelerate in rapidly so that other vehicles need to react accordingly. This creates a chain reaction leading to reduced travel speed. These are called *phantom jams* because they do not have any specific origin and are common in high density traffic regions like cities and high demand highways (ASTRA 2020).

This general layman's definition of jams is essentially correct, but not sufficient for the data scientific approach in this thesis. The Bavarian ministry for streets does not have an official definition at the time of writing and there is no unified definition or thresholds when reduced speed or time delays can classify data as congested or slow. This makes it necessary to form a specific definition of jams and their speed/space/time thresholds for the scope of this thesis. For instance the ADAC classifies highway traffic moving with mean speed lower than 20 km/h as jammed (ADAC 2019). In Switzerland the ministry for streets has a more severe definition with a mean speed of under 10 km/h (ASTRA 2020). A definition based just on the available literature does not consider the data is would be applied on. Therefore the speeds occurring in jams should be considered to further tailor the definition to our needs.

2.2.2 Data Scientific Definition

Figure 2.1 shows a section of a exemplary speed matrix plot from the FCD dataset (introduced in section 4.1) containing a scattered congestion cluster. The horizontal and vertical axis represent

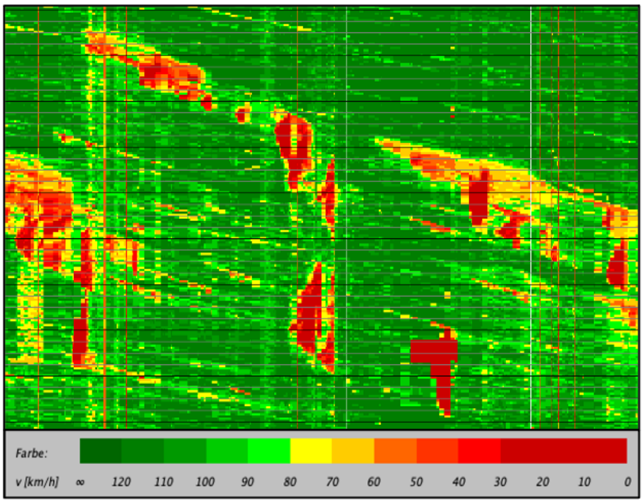


Figure 2.1: Speed matrix plots of FCD data, showing a scattered cluster

the spatial and temporal location of each cell. Each rectangle indicates the mean absolute speed recorded in the time frame and on the road section of the rectangle (a detailed gradient is shown in the legend of fig. 2.1). These rectangles, which correspond to a time frame of 3 min and a specific road section will be referred to as speed cell in the following.

The visual representation shows that cells of a congestion mostly contains speeds of less than 30 km/h, shown in *dark red*. When considering the cluster on the left, speed around 40-70 km/h, which are visualized in *lighter red* and *orange* tones, may be also relevant, to incorporate the complete congestion area. Speeds above at least 70 km/h, starting with the *yellow* categories, should not be considered, because it would include much noise. This makes two speed thresholds for jammed and slow-moving traffic necessary to adequately detecting congestion clusters. With the finding from the previous visual interpretation and some learning during implementing and calibrating the clustering algorithm (see section 5.1.1) the following thresholds for the jammed and slow speed, classifying jams in FCD were defined.

- Speed threshold for jammed state : $v_{\text{crit,jammed}} = 30\,\frac{\text{km}}{\text{h}}$
- Speed threshold for slow-moving state : $v_{\text{crit,slow}} = 60\,\frac{\text{km}}{\text{h}}$

To exclude cell errors and discard detections, too small to be considered as jams, the length and duration is used for filtering.

- Minimum length of a congestion : $l_{\text{min}} = 1000\,\text{m}$
- Minimum duration of a congestion : $t_{\text{min}} = 9\,\text{min}$

If $l < l_{\text{min}}$ or $t < t_{\text{min}}$ is given, l being the maximum spatial extend and t being the maximum temporal extend, the detection should be ignored.

3 Applied Statistic Methods

This chapter will explain the implemented statistical methods and their mathematical definitions. This should result in a deep understanding how the results which are presented in the upcoming chapters are comprised. It also clarifies which methods and assumptions are implemented to handle a mixed data-based analysis.

Correlation is an analysis procedure that measures the correlation coefficient, which represents the degree of linear, bivariant, monotonic or other kind of relation, which could also be described as the degree of association between two variables (Herz et al. 1992). In most statistics the following common types can be found: Pearson's r, Kendall's τ, Spearman ρ or the Point-Biserial correlation (Ramzai 2020; SPSS 2020a,b). Besides off these, there are many more correlation coefficients which vary in their applicability and interpretability. Depending on which type of data variables are to be analyzed, it is necessary to choose an applicable correlation coefficient. The type of data variable and relation combination are the most restricting features for choosing a suitable correlation coefficient.

3.1 Variable Types

Data variables can be grouped into continuous and categorical variables, depending on what kind of observation they describe. Variables are considered to be continuous, also known as quantitative, if they relate to measurements like speed, distance or age, which can take on an unlimited number of values between the lowest and highest points of measurement (McCue 2007). These continuous variables can be separated into two subsets.

- **Interval** variables can be measured along a continuum and have a numerical value (Laerd 2020).
- **Ratio** variables are interval variables, with the added condition that the values are set to zero if there is no measurement for this value (ibid.).

Categorical variables on the other hand are limited in the number of values, referring to a category, rank or choice, like a vehicle type or Yes/No answers. These categorical variables can be separated into three subsets.

- **Nominal** variables have two or more categories, but with no intrinsic order (ibid.).

- **Dichotomous** are nominal variables which have only two categories or levels (Laerd 2020).
- **Ordinal** are nominal variables that have two or more categories and are ordered or ranked (ibid.).

The datasets to be examined in this thesis include continuous variables of the type interval and all three types of categorical variables (see sections 4.2 and 4.3).

3.2 Correlation Types

The datasets from BAYSIS, ArbIS (see sections 4.2 and 4.3) and the processing tool (see section 5.3) includes continuous, as well as categorical variables, describing interval, nominal, dichotomous and ordinal characteristics. For an exact analysis of relations between these characteristics the appropriate correlation coefficient suited for the respective variables needs to be chosen. This task itself is quite complex due to the number of coefficients to evaluate, amount of literature and numerous assumptions in the field of statistics. From a comprehensive literature review the following correlation coefficients and tests were selected because of their appearance in current studies and papers and in dependency of their suitability for the respective relation.

For a fundamental introduction into correlation statistic, the article of Jun Ye *Everything you need to know about correlation*[1] is highly recommended (Jun 2020).

3.2.1 Correlation coefficient for continuous - continuous relations

As stated in section 3.1 continuous variables are metric measurements in form of distances or durations. The most common correlation coefficient for continuous variables is the so called non-parametric Pearson's r.

Pearson's r

Pearson's r describes the linear correlation of continuous, non-ranked variables and does not assume normality or a normal distributed sample set as it is non-parametric. It is therefore suitable for the examination of continuous - continuous variable relations, where finite size of variance and covariance can be assumed. (Benesty et al. 2009; Sulthan 2018)

The general correlation coefficient, shown in eq. (3.1) is the foundation for deducing Pearson's r. It is defined by the fraction of the covariance σ_{xy} of two vectors x,y of length i, see eq. (3.2) and their standard deviation $\sigma_{x,y}$, see eq. (3.3) (Herz et al. 1992).

[1] https://junye0798.com/post/everythin-you-need-to-know-about-correlation/

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{3.1}$$

$$\sigma_{xy} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}} \tag{3.2}$$

$$\begin{aligned} \sigma_x &= \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n}} \\ \sigma_y &= \sqrt{\frac{\sum_i^n (y_i - \bar{y})^2}{n}} \end{aligned} \tag{3.3}$$

The following eq. (3.4) shows Pearson's correlation coefficient r, which is a direct usage of the definition in eq. (3.1), assuming that both data variables have the same length, named i. The symbols $\bar{x}$ and $\bar{y}$ correspond to the means of the data variable x and y, respectively. (Benesty et al. 2009; Zychlinski 2018)

$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{3.4}$$

This equation can be simplified to eq. (3.5) with the SS corresponding to the summed squares and SP corresponding to summed products.

$$r = \frac{SP_{xy}}{\sqrt{SS_x SS_y}} \tag{3.5}$$

Interpretation of r

Pearson's r can have values of the range -1 to $+1$. If one variable moves in the same direction as the other, it is called positive correlation, represented by a positive correlation coefficient. In the case of one variable changing in a positive direction, whereas a second variable is changing in a negative direction, the correlation is called negative and has a negative coefficient. Another characteristic is the rate of change in the variables. When both variables change at the same rate, they are linearly correlated. When both variables do not change in the same rate, then they are non-linearly or curvy-linear correlated and r approximate to zero, showing no correlation.

The effect size of Person's r defined the absolute value of r, mathematic written as $|r|$. According to Cohen recommendations for mean-based coefficients (explained in section 3.4) and with consideration of the guidelines by Wolfe (Wolfe et al. 2017) and Regber (Regber 2016), which advocate an increased scale due to the high sensitivity of Pearson's r, the following rules are defined for the interpretation of the effect size of r.

- When both variables change in the same ratio, the absolute value is 1.0, which is called perfect correlation.
- If the range is above .80, it is called high degree of correlation.
- A moderate degree of correlation lays in the range of .50 to .80.
- When the range is between .30 to .50, it is called low degree of correlation.
- If the range is lower than .30 no correlation can be proven. It is called absence of correlation.

Significance of r

To determine if r is statically significant a chi-square test is generally applied to find the p-value, testing the probability of independence (see section 3.5.2).

3.2.2 Correlation coefficient for continuous - nominal relations

This type of relation is objectively the most complex to evaluate. One well known method to analyze the relation between a continuous and categorical variables, which is not ranked and has more than two values, is the analysis of variance (ANOVA). Unfortunately, it assumes normal or gaussian distributed variables, which are not given in the dataset (see chapter 4), as it is categorized as a parametric test. A non parametric approach of the ANOVA is the rather uncommon Kruskal-Wallis h-test (Leon 1998). Both tests indicate if at least one variable stochastically dominates another, but not in which groups or in how many groups this domination occurs (Outside Two Standard Deviations 2020). They therefore do not provide a statement about the correlation strength, but the statistical significance of variances between groups.

Further research of a non-parametric correlation coefficient only provided one suitable option, the eta (η) coefficient by Pearson (Benninghaus 2007), describing the relationship between variables based on the sums of squares used in the ANOVA (Benninghaus 2007; Sir Arthur Lewis Institute 2012).

Eta (η) coefficient

The η coefficient, also called correlation ratio, is a measurement for the proportion of the variation in y, which is associated with a membership of different groups in x (Lakens 2013). Or mathematically η is the squared root of the ratio between SS_x and SS_y (Hedayatpanah Shaldehi 2013; SAGE 2014). When calculated the value of η^2 represents the percentage of total variance which can be accounted to a group relation (Lakens 2013).

$$\eta = \sqrt{\frac{SS_x}{SS_y}} \tag{3.6}$$

Interpretation of η

The coefficient η can have values in the range from 0 to 1 and can be interpreted similar to Person's r according to (ibid.). The effect size, which is categorized in the R group and can therefore be interpreted as follows (Cohen 1988; Regber 2016):

- $\eta <= .06$: the correlation strength is weak
- $.06 < \eta < .14$: a moderate strength of correlation
- $\eta >= .14$: there is a strong correlation.

Significance of η

The significance evaluation for the η coefficient is generally performed via the p-value of the Kruskal-Wallis h-test, since it is a non-parametric test. It is calculated with the chi-squared test (χ^2), defined in section 3.5.2 (Filipiak et al. 2013). In the case of both a correlated η and a significant h-test, the exact related groups need to be determined via a post-hoc test (see section 3.6).

3.2.3 Correlation coefficient for continuous - dichotomous relations

The Point Biserial correlation is a special form of the Pearson's r correlation coefficient and suited to evaluate the association of continuous-dichotomous relations.

Point Biserial

The Point Biserial notation, shown in eq. (3.9), can be derived from Person's r with the assumption of y only taking dichotomy values of 0 and 1, so that $\bar{y} = p$. The distinction of the cases

- $n \cdot p$ referring to $y = 1$ an with $1 - p = q$ bigger than $\bar{y}$
- $n \cdot q$ referring to $y = 0$ an with $0 - p = -p$ smaller than $\bar{y}$

allow to form eq. (3.7) from eq. (3.4), which can be simplified to eq. (3.9) with the intermediate step of eq. (3.8) (Bortz 2004; Cohen et al. 2003; DeJesus 2019; Tate 1954).

$$r_{pq} = \frac{n \cdot p(\bar{x}_{y=1} - \bar{x}) \cdot q + n \cdot p(\bar{x}_{y=0} - \bar{x}) \cdot (-q)}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot (n \cdot p \cdot q^2 + n \cdot q \cdot (-p)^2)}} \tag{3.7}$$

$$r_{pqi} = \frac{n \cdot p \cdot q \cdot (\bar{x}_{y=1} - \bar{x}_{y=0})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot (n \cdot p \cdot q)}} \tag{3.8}$$

$$r_{pq} = \frac{\bar{x}_{y=1} - \bar{x}_{y=0}}{\sqrt{\sum_i (x_i - \bar{x})^2}} \cdot \sqrt{n \cdot p \cdot q \cdot} \tag{3.9}$$

It must be pointed out that if the dichotomous variable is artificially binarized, i.e. there is likely continuous data underlying it, biserial correlation is more a measurement of similarity instead of association (Outside Two Standard Deviations 2020).

Interpretation of r_{pq}

Due to the mathematical similarity of the Point Biserial to the Pearson's r, the general interpretation of Pearson's r defined in section 3.2.1 can be applied to Point Biserial with some adjustments. The range of the Point Biserial coefficient, from 0 to 1 removes the direction of correlation from the interpretation. According to Cohen (Cohen 1988) the following can be used as guidelines for the effect size r_{pq} (Leblanc et al. 2017):

- $r_{pq} < .30$: The correlation strength is weak
- $.30 =< r_{pq} < .50$: a moderate strength of correlation
- $r_{pq} >= .50$: there is a strong correlation.

Significance of r_{pq}

The significance evaluation of the Point Biserial coefficient is done via the 2-tailed p-value, a doubled chi-square test (see section 3.5.2).

3.2.4 Correlation coefficient for continuous - ordinal relations

For ordinal variables, also called ranked or rank ordered, the commonly used Spearman's ρ can be applied, but should be replaced by Kendall's τ because of it superiority over Spearman (Newson et al. 2002).

Kendall's τ

Kendalls τ evaluates the order of rank pairs, instead of the squared rank difference, which makes it more robust against outliers. Because it can be assumed that the data has ties implemented, the τ with ties must be used. The general definition is shown in eq. (3.10), with P referring to the proversion and I to the inversion. With the assumption that the continuous measurement x is or can be ordered, the ordinal/ranked variable y will be wrongly ordered. After forming all possible rank pairs between x and y, P and I can be deduced and τ can be calculated. (Brossart et al. 2017; Reiter 2015)

Proversion $(+)$ is the number if pairs, where $x < y$

Inversion $(-)$ is the number of pairs, where $x > y$

Ties (0) are pairs where $x = y$

$$\tau = \frac{P - I}{\sqrt{\left(\frac{N(N-1)}{2} - T_x\right) \cdot \left(\frac{N(N-1)}{2} - T_y\right)}} \tag{3.10}$$

$$T_x = \sum_{i=1}^{n} \frac{t_{x_i}(t_{x_i} - 1)}{2} \tag{3.11}$$

$$T_y = \sum_{j=1}^{m} \frac{t_{y_j}(t_{y_j} - 1)}{2} \tag{3.12}$$

N is the total number of rank pairs

$P|I$ are the pro- and inversion of pairs

$T_x|T_y$ are the ties in x and y

$n|m$ are the number of rank bindings in x and y

$t_{x_i}|t_{x_i}$ are the length of rank bindings in x and y

Through transformation of eq. (3.10) we can simplify the general definition to equation 3.13 (Reiter 2015).

$$\tau = \frac{P - I}{\sqrt{(P + I + T_x) \cdot (P + I + T_y)}} \tag{3.13}$$

Interpretation of τ

Strictly speaking, Kendall's τ is not a measure of effect size, like Pearson's r, but tends to be of similar magnitude. Because of this similarity the general interpretation defined in section 3.2.1 and section 3.4 can be applied. To adapt the guidelines to the lesser sensitivity of τ, they are scaled downwards (Regber 2016).

- $\tau <\ .30:$ The correlation strength is weak
- $.30 =< \tau <\ .50:$ A moderate strength of correlation
- $\tau >=\ .50:$ There is a strong correlation

Significance

The significance evaluation of Kendall's τ coefficient is carried out via the 2-tailed p-value, elaborated in section 3.5.2.

3.2.5 Correlation coefficient for categorical - categorical relations

The Pearson's χ^2 test from section 3.2.1 can also be applied to categorical data for independence statistics. Two correlation coefficients using χ^2 are Cramer's V and Theil's U. Both can be used to analyze categorical - categorical relations, but differ in the type of result they provide (Outside Two Standard Deviations 2018). Cramer's V is a symmetric measure providing a measure of association strength. Theil's U, the uncertainty coefficient, on the other hand is a conditional measure and represents the predictability of an association (Akoglu 2018; StackExchange 2020). Therefore the Theil's U measurement (see section 3.7) it is the preferred choice for the analysis of predictability and Cramer's V is used in the correlation analysis.

Cramer's V

Cramer's V, also called Cramer's phi (Φ_c), is a measurement for the relation of two nominal variables. In eq. (3.14), showing the notation of Cramer's V, k and r are the number of columns and rows, respectively. φ, the phi coefficient, is defined by $\frac{\chi^2}{n_{ij}}$. The χ^2 shown in eq. (3.15) is derived from eq. (3.21) with the expansion to columns and rows (Bergsma 2013; J. Sheskin 1997).

$$V = \Phi_c = \sqrt{\frac{\varphi^2}{min(k-1, r-1)}} = \sqrt{\frac{\frac{\chi^2}{n_{ij}}}{min(k-1, r-1)}} \tag{3.14}$$

$$\chi^2 = \sum_{i,j} \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{\frac{n_i n_j}{n}} \tag{3.15}$$

The above notation of Φ_c can be heavily biased, trending to overestimate the strength of relation. It can be corrected with eq. (3.16), using the corrected notation eq. (3.17) for $\tilde{\varphi}^2$ and eq. (3.18) as well as eq. (3.18) for k, r. (Bergsma 2013)

$$\tilde{V} = \tilde{\Phi}_c = \sqrt{\frac{\tilde{\varphi}^2}{\min(\tilde{i}_{\max} - 1, \tilde{j}_{\max} - 1)}} \tag{3.16}$$

$$\tilde{\varphi}^2 = \max\left(0, \varphi^2 - \frac{(k-1)(r-1)}{n-1}\right) \tag{3.17}$$

$$\tilde{k} = k - \frac{(k-1)^2}{n-1} \tag{3.18}$$

$$\tilde{r} = r - \frac{(r-1)^2}{n-1} \tag{3.19}$$

Interpretation

For the interpretation of Cramers V an adaption to Pearson's r is necessary, but at the same time quite controversial. Some studies convert V to the effect size w for an equal measurement to r (r is representable over different studies), whereas V is already a measure of effect size by itself (Baguley et al. 2016).

$$w = V \cdot \sqrt{\min(i_{\max} - 1, j_{\max} - 1)} \tag{3.20}$$

As shown in eq. (3.20) (ibid.) the conversion from V to w is similar to the reduction of Φ_c in V (see eq. (3.16)). This conversion to w is necessary within a single study and an adaption of scale is sufficient (ibid.). In terms of this thesis an adapted scale of effect size according to Ellis is used to interpret w in the range of 0 to 1 (Cohen 1988; Ellis 2010; Hemmerich 2019)

- $w <\ .30:$ the correlation strength is weak
- $.30 =< w <\ .40:$ a moderate strength of correlation
- $w >=\ .40:$ there is a strong correlation.

3.3 Correlation Coefficient Matrix

As a result the following correlation coefficients and statistical tests are used for the mixed analysis of continuous and categorical variables. These will be implemented into a correlation processing script, which is explained in section 5.4.

	Ordinal	**Dichotomous**	**Nominal**	**Continuous**
Ordinal	Cramer's V	Cramer's V	Cramer's V	Kendall's τ
Dichotomous	Cramer's V	Cramer's V	Cramer's V	Point Biserial r
Nominal	Cramer's V	Cramer's V	Cramer's V	Eta η
Continuous	Kendall's τ	Point Biserial r	Eta η	Pearson's r

Table 3.1: Selected Correlation Coefficients

3.4 The Effect Size

The degree of correlation, also described as the strength of association is called effect size and shows how strong two variables are related with each other. According to Cohen (Cohen 1988) the effect size of each correlation coefficient falls into one of two categories D and R. D corresponds to coefficients utilizing the mean difference and standardized mean difference. He defined the values of D coefficients as small $D = .20$, medium $D = .50$, and large $D = .80$ (Piegorsch 2002). The group of R coefficients includes measures based on variance (Walker 2005). Cohen proposes vastly different values of .01, .06, and .14 as indicators of small, medium, and large effect sizes for the R group (Cohen 1988). However, if these values fit the purpose of the analysis, depends on the underlying data and is at discretion of the researcher. This means the values for interpreting the effect size can and should be adapted to fit the underlying data and the usage of the interpretation.

There are many effect sizes like Cohen's f^2 or Heghes g, which can be used for the interpretation of relations. The interpretation by Cohen for r is the most robust and common ground found in literature and will be used as a mathematical base for other correlation coefficients.

3.5 Significance vs. Uncertainty

Pure correlation results are not sufficient to prove a relation between variables, especially when using the correlation evidence for predictive purposes. Depending on the sample data, statistical results can be random or biased and therefore need to be examined for their probability of error. This is commonly evaluated via the well known statistical significance. Even though statistical significance is the common standard procedure for evaluating the probability of error, it became a subject of debate in recent years if statistical significance is to be generally used and if it is potentially misleading. It is often advocated to additionally used statistical uncertainty instead of only relying on statistical significance (Harris 2019).

3.5.1 Uncertainty

The uncertainty assessment aims to determine if a variable has suitable statistical characteristics for the intended analysis or in colloquial terms, is *fit for the purpose.* The literature research showed that there are a number of different procedures for the evaluation of uncertainty. *The Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement*[2] (also referred to as GUM) defines reliable but also complex guidelines and is recommended for an in depth understanding of uncertainty evaluation (Farrance et al. 2012). The complexity of the GUM procedures (which are beyond of the scope of this thesis) and the

[2] https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

data foundation favored the usage of simpler methods like the *population sampling error* (ONS 2020).

The sampling error is the uncertainty created by selecting a subset from a population. This can result in poorly distributed variables, which induces a high probability of random correlations based on a small number of samples. An initial review of the variable distributions in the BAYSIS and ArbIS dataset showed that there are many sample sizes below 10 (see sections 4.2 and 4.3), which is an arguably low count. Research into the minimal sample size showed, that for categorical variables, with number of categories k a minimal number of 20 ($k = 3$) or 200 ($k = 10$) should be asserted (Cicchetti 1981). From these guidelines and the exploratory approach of this thesis, which allows for lower thresholds, a sample size below 10 are considered to be uncertain and associated correlations should be neglected by default.

3.5.2 Significance

Because statistical results are usually based on sample sets, it is necessary to test if the results can be applied on the population or are significant. Therefore a decision on one of the two hypothesis of *"The averages are equal to the population"* which means the statistical results can be applied on the general population and *"The averages are not equal to the population"* which means the statistical results can not be applied on the general population, need to be determined. This can be evaluated by the common dependent t-test which compares the significance value p to the error probability level α. The p-value is generally calculated via chi-squared, shown in eq. (3.21) (chi-squared statistic), which was taken from Wikipedia and cited from Karl Pearson (Pearson 1900).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^{n} \frac{(O_i/N - p_i)^2}{p_i} \tag{3.21}$$

N is the total number of data samples

O_i is the number of data samples with type i

$E_i = Np_i$ is the expected number of data samples with type i

The p-value can then be comprised by comparing χ^2 to a χ^2-distribution by calculation or by using a conversion table (Piegorsch 2002) with the degree of freedom $df = (n_x - 1) \cdot (n_y - 1)$. The resulting p-value is compared to the α-level to either accept ($p > \alpha$) or reject ($p \leq \alpha$) the null hypothesis *"The means are equal to the population"*. For a 2-tailed p-value, p will be doubled to incorporate both ends of the distribution. Usually a value of .05 is chosen for α which there will be a 5 % risk of falsely rejecting the null hypothesis. Out of this definition the following two interpretations of the correlation coefficient can be drawn (Outside Two Standard Deviations 2020; Tenny et al. 2020).

- $p <= \alpha$ means that the null hypothesis can be rejected and indicates that there is a significant dependency between the two tested variables. It can be concluded that the increase or decrease of one variable does significantly relate to the increase or decrease of the other.
- $p > \alpha$ means that there is **no** significant dependency between the two variables and no conclusion can be drawn from the correlation.

3.6 The Post Hoc Analysis

In the case of a relation of a categorical variable the correlation effect size and its significance do not provide any statement on which categories are related. This needs to be further evaluated with a Post Hoc test. In our case the Post Hoc test is based on two separated tests. First the variables will be tested for significance with a Kruskal-Wallis rank sum test.

$$h = \frac{12}{N(N+1)} \sum_h \frac{S_h^2}{n_h} - 3(N+1) \tag{3.22}$$

N is the total number of samples

n is the number of samples in group h

S_h rank sum of group h

For Kruskal-Wallis h-test the following rules apply. The higher the value of h, the higher is the variance between the unique variables. A high h normally supports an already found correlation, but must be tested for significance with χ^2 defined in section 3.5.2. If the significance or p-value is below the α-level of .05 the null hypothesis can be rejected. This implies that the correlation or variances between the groups are significant. For the identification of which groups create this significant variance the following Wilcoxon-Mann-Whitney, sometimes also called Wilcoxon T-test is applied pairwise.

The Wilcoxon T-test is a non-parametric univariate alternative to the dependent t-test and is the recommended test to use when the data violates the assumption of a normal distribution. The two variations *Mann-Whitney U statistic* and Wilcoxon W sanked sum statistic can be considered as equivalent. The Wilcoxon ranked sum test will be used in a pairwise approach to compare all groups with each other.

From the resulting matrix of p-values for each group ($G_{1,\ldots n}$) and the analog interpretation with the α-level (see section 3.5.2) the significant groups can be identified. These identified groups are considered to have a significant difference to the other groups and therefore can predictably point to a measurement characteristic from a category or vice versa.

$$
\begin{tikzcd}
p(G_1, G_1) \arrow[r] \arrow[d, "{1, ..., n}"'] \arrow[dr, dashed, no head, "0", "p"'] & p(G_n, G_1) \arrow[d] \\
p(G_1, G_n) \arrow[r, "{1, ..., n}"'] & p(G_n, G_n)
\end{tikzcd}
$$

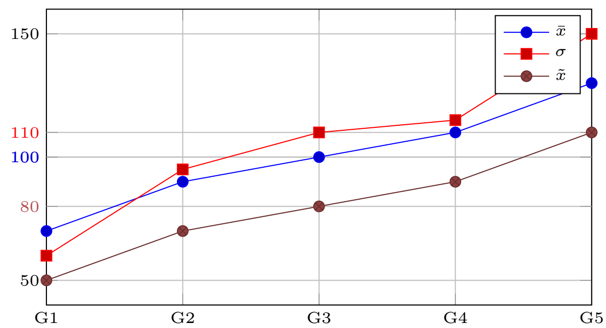
Figure 3.1: 2-dimensional Wilcoxon ranked sum test significance matrix

A non significant difference does not mean that there is no correlation, but only that the sample set and size can not support significance. There are also cases where the p-value of the correlation coefficient and the general Kruskal-Wallis rank sum test show significant differences in the variable, but the pairwise Wilcoxon ranked sum test does not show any significant differences between the groups. This unfitting difference between the *global* and *specific* significance appears when variables have differences between the samples, but the differences are not reflected by the groups or the sample size is not sufficient to support the significance of the differences. These controversial correlations can also be evaluated in a exploratory data analysis, keeping in mind their limited interpretability and predictability.

To define what the significant, as well as the non significant groups are referring to in the variable, the statical indicators of count (n), mean ($\bar{x}$), standard deviation (σ), median ($\tilde{x}$), min, max and range (Δ) are considered. These are quite common for describing statical data and can be considered as robust. These indicators can be presented as a table or visual plot, as shown in figs. 3.2a and 3.2b.

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
G_1	n_1	$\bar{x}_1$	σ_1	$\tilde{x}_1$	min_1	max_1	Δ_1
...	...	...	...	...	...	...	...
G_{n-1}	n_{n-1}	$\bar{x}_{n-1}$	σ_{n-1}	$\tilde{x}_{n-1}$	min_{n-1}	max_{n-1}	Δ_{n-1}
G_n	n_n	$\bar{x}_n$	σ_n	$\tilde{x}_n$	min_n	max_n	Δ_n

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 3.2: Group descriptives example

The example table shows the values of all descriptives for all groups, whereas the example figure on the right show a plot of mean ($\bar{x}$), standard deviation (σ) and median ($\tilde{x}$) with the color coded means. The figure also shows the means of the variables $\bar{x}$, σ and $\tilde{x}$ in there corresponding color code. From these the absolute value of difference between the groups can be deviated by looking for outliners or distinct variances in the distributions, which can be interpreted accordingly. The indicators of min, max and range (Δ) can give clues what distributions should be expected and if the variable is sufficiently distributed.

3.7 The Predictability Analysis

The uncertainty coefficient Theil's U, also called entropy coefficient, is a measurement for the association between two nominal variables and in comparison to Cramers V, provides a much better predictability statement. It is based on the concept of comparing the entropies of variables to determine a degree of association (Hoang 2019). The entropy of a distribution (see eq. (3.24)) and the conditional entropy (see eq. (3.25)) are used to calculate the uncertainty coefficient $U(X)$ (Glen 2017, 2018), which tells us: given Y, what fraction can be predicted for X (Hoang 2019).

$$U(X) = \frac{H(X) - H(X|Y)}{H(X)} \tag{3.23}$$

$$H(X) = -\sum_{x} p_X \log_{p_X}(x) \tag{3.24}$$

$$H(X|Y) = -\sum_{x,y} p_{X,Y} \log_{p_{X,Y}}(x, y) \tag{3.25}$$

The value U of Theil's U shows the predictability of the association, where the following rules for interpretation apply to the implementation (Ahlburg 1984; Bliemel 1973; Granger et al. 1973).

- U close to 0 : y provides no information about x
- U close to 1 : y provides full information about x

4 Data Foundation

In this chapter the provided datasets from the previously mentioned street information systems and the FCD provider will be elaborated. In the beginning an overview of FCD in general and the available dataset is given. To give an overview about which results can be expected, the incident datasets from BAYSIS and ArbIS are presented with descriptive statistic. The most relevant parameters of the datasets will be elaborated and illustrated. The parameters are also categorized in the variable types defined in section 3.1.

4.1 Floating-Car-Data (FCD)

As described in section 1.1, FCD represents the movement of vehicles and can be used to calculate vehicle speeds and trajectories. The provided dataset contains the aggregated absolute and relative speeds for highways and state streets, calculated from FCD data. The process of speeds estimation with FCD data is explained in detail by Felix Rampe in chapter 4 of his thesis *Traffic Speed Estimation and Prediction Using Floating Car Data* (Rempe 2018), but goes beyond the scope of this thesis. The FCD is then mapped onto the HERE (HERE 2020) network, to be compliant with the geolocation system used in the project. Each of these aggregated speeds now represents the mean speed over a three-minute time interval on the corresponding road section. This arrangement of speeds for each time step and space step can be called speed matrix and is the base data for the congestion detection. Figure 4.1 shows a visual representation a speed matrixes with the horizontal axes being the spatial extend and the vertical axes the time extend.

Deep greens represent free flowing traffic with 130 km/h, according to the norm speed on highways set the legislator in Germany (in german called "Richtgeschwindigkeit"). The speed scale then develops linearly downwards to deep red indicating traffic with 30 km/h or less. The observer will clearly recognize the jams represented by the clusters of red and orange cells in the fig. 4.1. The vehicle trajectory through space and time can be recognize on the angled extends from the top left to the bottom right. From this visual clarity speed matrix and the precision on 3-minute intervals it can be concluded that a comprehensive algorithmic approach should be able to detect such congestion events. This being said, the dataset does contain defects in the form of missing values for complete road sections, which can interfere with detection algorithm. Another defect type would be an obviously wrong speed block, meaning sudden speed drops or jumps to areas of

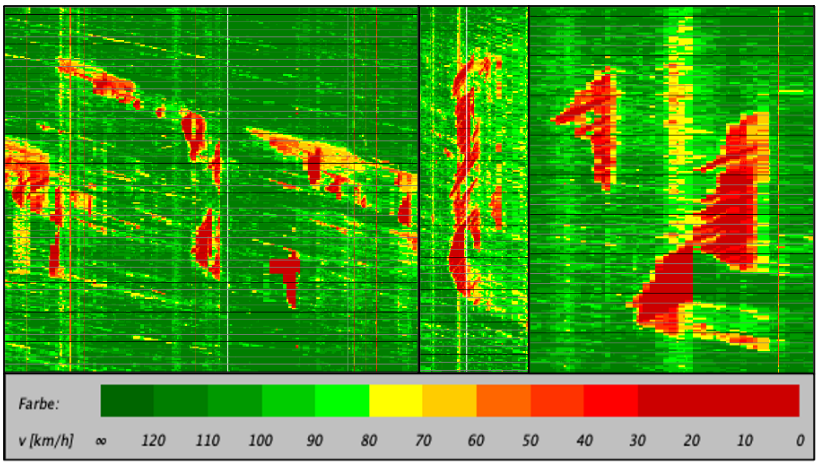


Figure 4.1: Speed matrix plots of processed FCD data, showing different jam clusters

identical speeds with an abnormal temporal and spatial extend, which then have to be ignored during processing.

4.2 Accident Data (BAYSIS)

The Bayerische Straßeninformationssystem (BAYSIS) as described in section 1.2, collects a wide range of different information types, one of them being accidents with the corresponding police reports. The provided export from BAYSIS contains all accidents of the year 2019 on the Bavarian highway network, which are 10262 records in number. Each accident report includes a variety of specifications which covers environmental indicators like weather or light situation, accident characteristics like accident type, collision object or cause, as well as information over the involved persons like nationality, age and gender. In total, one report contains 132 values describing the accident, participant and environment. As there shouldn't be formed a stereotype of accident participant but rather significant accident characteristics or environmental factors be found, most of the descriptive values for the involved persons are not considered. Variables which contain no values or a single values and fall under the uncertainty threshold (see section 3.5.1) are also neglected. From this curtailed pool of correlate able and analyze able characteristics all parameters that have a logical significance with causes or effects of an accident will be considered in the analysis. The referred figures are available as larger prints in the appendix chapter A for better readability, as they do not properly fit in the text.

A look on the monthly distribution of accidents recorded by BAYSIS (see fig. 4.2) shows that that the months of January, July and August show a considerably high number of accidents, with respectively 31 %, 21 % and 15 % increase over the mean count of 855 accidents per month. The increased number in January can be explained with the increased number of accidents due to ice and snow conditions, which reduces traction on roads and can lead to uncontrollable vehicle behavior. Also the reduced visibility during snow falls increases accident numbers. In July and August the increased traffic volume because of public holidays is the most probable explanation

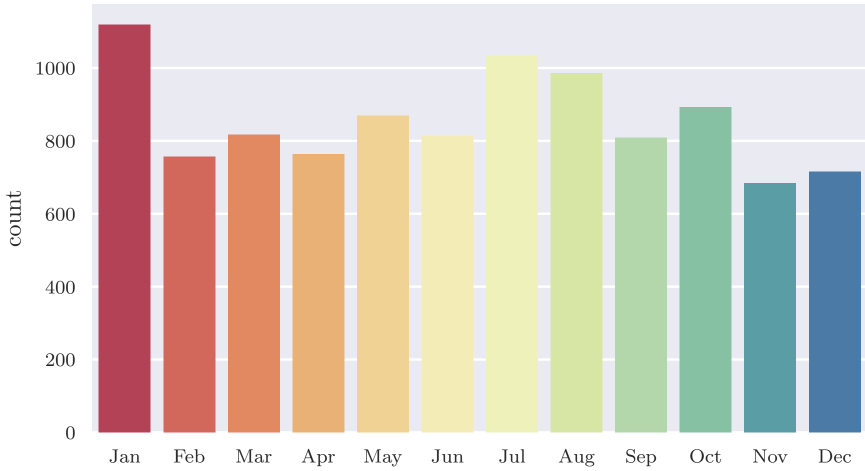


Figure 4.2: Distribution of accident counts by months

for the higher number of accidents. Another valuable distribution is the number of accidents per road (see fig. 4.3). The roads A3, A9 and A8 show a relative high count of accidents whereas the road A71 and beneath only have a small number accidents.

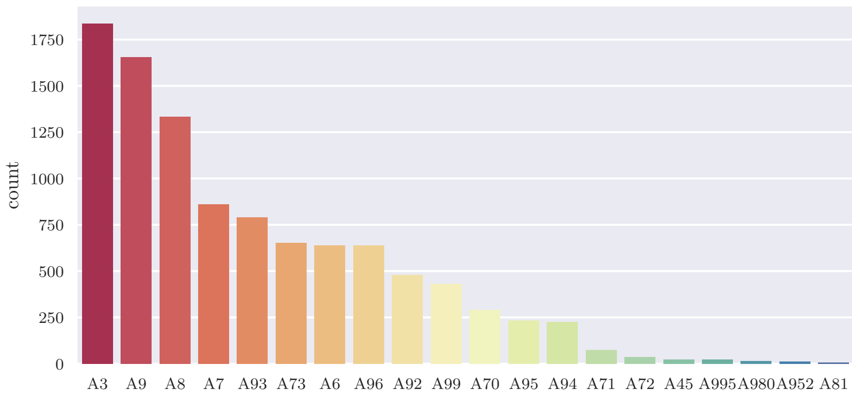


Figure 4.3: Distribution of accident counts by highway

Kat The accident category (shown in table 4.1 and visualized in fig. A.1), describes which damages or injuries can be associated with the accident. It ranges from accidents with just damaged property to lightly and heavily injured and even deathly accidents. The distribution develops from lowest to highest counts, in order of gravity of the category. The variable consists of four values, which can be ordered and is therefore ordinal.

Typ The accident type variable (shown in table 4.2 and visualized in fig. A.1) incorporates different kinds of traffic movements from straight driving to turning movements or merging. It describes during which kind of movement the accident happened. Beside of an 80 % share

Code	**Kat**	[%]	Description
0	-	-	Minor Accident
1	76	0.7	Accident with deaths
2	600	5.8	Accident with heavily injured
3	2685	26.2	Accident with lightly injured
7	6900	67.2	Accident with property damage

Table 4.1: Descriptive of **Kat**

of accidents related to driving or straight driving situations, the parameter does not indicate any other features. The variable does not show any order and is therefore of a nominal type.

Code	**Typ**	[%]	Description
1	2820	27.5	Driving accident
3	373	3.6	Merging / Crossing accident
4	10	0.1	Crossing over accident
5	160	1.6	Accident in standing traffic
6	5285	51.5	Accident in straight traffic
7	1612	15.7	Other

Table 4.2: Descriptive of **Typ**

Betei The distribution of the number of involved persons (shown in table 4.3 and visualized in fig. A.1) shows that more than 97 % of all accidents have three or less involved persons, which is supported by a mean of 1.9 involved persons per accident in general. The major share of two involved persons makes up for 59 % and the second biggest of one involved person for 30 % of the total count. Because of the increasing order of values, the variable is of an ordinal type.

	1	2	3	4	5	6	> 7
Betei	2768	6047	1085	235	72	24	30
Freq.	27.0	58.9	10.6	2.3	0.7	0.2	0.1

Table 4.3: Descriptive of **Betei**

UArt The accident cause type is defined by the two variables **UArt1** and **UArt2** (shown in table 4.4 and visualized in fig. A.1), whereat the first is the more suitable one. They describe the type of collision cause and presents two major sets. One being the accidents with waiting, stopping and starting vehicles in the same lane, which describe typical collision accidents during congested traffic. The other being the accidents in the next left or right lane, which describe common lane changing collisions. Accidents with cross traffic, pedestrians or opposite traffic are relatively uncommon. The variable does not show any order and is therefore of a nominal type.

Code	**UArt1**	**UArt2**	Freq.	Description
1	466	13	4.2	Collision with starting, standing or stopping vehicle
2	2111	44	18.8	Collision with ahead and waiting vehicle
3	2873	140	26.3	Collision with vehicle on separate lane in same direction
4	16	4	0.2	Collision with vehicle going in opposite direction
5	240	12	2.2	Collision with turning or crossing vehicle
6	19	1	0.2	Collision between vehicle and pedestrian
7	411	42	4.0	Collision with obstacle
8	1728	381	18.4	Deviation to the right
9	1446	517	17.1	Deviation to the left
0	951	42	8.7	Other

Table 4.4: Descriptive of **UArt**

AUrs The summarized distribution of the parameters **AUrs1** and **AUrs2** (shown in table 4.5 and visualized in fig. A.2) shows that the variable is not much distributed and only a small number of categories hold a relevant sample size. Any correlation to this parameter needs to interpreted with caution, due to the high uncertainty. The variable is of a nominal type.

Code	**AUrs1**	**AUrs2**	[%]	Description
72	618	-	33	Slippery street due to snow or ice
73	855	45	48	Slippery street due to rain
75	11	12	1.2	Cart track due to rain, snow or ice
76	10	3	0.7	Other condition of road
81	4	17	1.1	Visibility issues due to rain or hail
82	24	-	1.3	Visibility issues due to sun or glare
84	1	14	0.8	Visibility issues due to storm
86	21	-	1.1	Wild animals
88	134	4	7.4	Other obstacles
89	97	4	5.4	Other causes

Table 4.5: Descriptive of **AUrs**

AufHi The obstacle collision distribution (shown in table 4.6 and visualized in fig. A.2) reveals that in most collision accidents cars hit the guardrails. The other categories are rather uncommon. With 1,4 % of accidents without any collision, it can be stated that in most cases a collision is part of an accident. The counts of the remaining categories are insignificant. The variable does not show any order and is therefore of a nominal type.

Code	**AufHi**	[%]	Description
0	16	0.2	Single tree
1	12	0.1	Pillar
3	3041	29.6	Guardrail
4	534	5.2	Other object
5	144	1.4	No collision
8	26	0.3	Tree group or forest
9	52	0.5	Busches

Table 4.6: Descriptive of **AufHi**

Alkoh The alcohol involvement indication variable only contains one variables of (1 = *yes*), whereas an empty variable referred to *no* or *unknown*. It reveals that only 2.2 % of accidents have one or more involved persons with measurable blood alcohol. The variable only has two unique values and is therefore dichotomous.

Char The variable Char1 and Char2 (shown in Table 4.7 and visualized in A.2) describes the characteristics of the street where the accident happened. Since only highway are considered, the type of *Crossing*, *Property* and *Roundabout* is expected to be zero. The variable is not ordered and therefore of a nominal type.

Code	**Char1**	**Char1**	[%]	Description
2	42	-	2.9	Entry / Exit
4	380	-	26	Incline
5	371	-	25.4	Decline
6	404	263	45.7	Curve

Table 4.7: Descriptive of **Char**

Bes The variables **Bes1**, **Bes2** and **Bes3** further define the street characteristic mentioned above, although they only contain one variable, referring to the category *Roadwork*. As the variable is not distributed it is not suitable for correlation analysis, but can be used to validate the roadwork matching performance.

Lich The light situation variable (shown in table 4.8 and visualized in fig. A.3) describes the lighting condition at the time of the accident. The variable **Lich1** describes the nature light setting, when **Lich2** describes if the street light was working. Because **Lich1** can be ranked from best to worst lighting is of an ordinal type. **Lich2** only has two values and is therefore dichotomous.

Code	**Lich1**	**Lich2**	[%]	Description
0	6833	-	51.7	Daylight
1	627	-	4.7	Noon
2	2560	-	19.4	Darkness
3	-	3005	22.8	Street lighting working
4	-	182	1.4	Street lighting not working

Table 4.8: Descriptive of **Lich**

Zust The road condition parameter (shown in table 4.9 and visualized in fig. A.3) describes in which condition e.g. wet, dry, iced or slippery the road was at time of the accident. The condition can be ranked from best to worst and is therefore of an ordinal type.

Code	**Zust**	Count[2]	[%]	Description
0	6851	-	66.9	Dry
1	2606	-	25.4	Wet
2	582	207	7.7	Ice

Table 4.9: Descriptive of **Zust**

Fstf The variable references the lane on which the accident happened (shown in table 4.10 and visualized in fig. A.3). It names the number of lane from the right, the hard-shoulder or the wrong usage of a one-way street. It shows a ranked order of the lane number, but not with the two other types of hard-shoulder (S) and one-way street (F). It is therefore considered as nominal type.

Code	**Fstf**	[%]	Description
1	2821	27.5	first lane from the right
2	4274	41.7	second lane from the right
3	1582	15.4	third lane from the right
4	150	1.5	fourth lane from the right
5	26	0.3	firth lane from the right
S	247	2.4	on the hard-shoulder lane

Table 4.10: Descriptive of **Fstf**

WoTag The variable of WoTag relates to the day of the week the accident happened (shown in fig. A.3). It is debatable if week days can be ordered but for this analysis the parameter will be considered a nominal type.

FeiTag Only 157 accidents took place on a public holiday. This is not a feature itself but there could be a possible correlation to it, as the jams could be longer because of the increased traffic demand on holiday.

The designed evaluation tool utilizes a PostgreSQL database for its data storage. Therefore the BAYSIS data in form of Comma-seperated Values (CSV) needs to be processed and converted into SQL data entities. Also, the data entities for each accident need to be uniform and comparable with the street network and other entities like roadworks, which makes it necessary to process and map the accidents onto our street network. After the necessary processing and import into the database, 7971 records end up being converted and persisted, which is equivalent to 77,6 % of the total number of accidents. This 22,4 % of data loss can be explained due to the conversion of from the BAYSIS geo-system to the HERE network, which tries to find a corresponding street network location to the legacy location of the BAYSIS dataset. If it is not able to locate the position of the BYSIS dataset on our street network, the record is discarded.

4.3 Roadwork Data (ArbIS)

The Arbeitstellenintegrationsystem (ArbIS) as described in section 4.3 is a collection service of all roadworks or maintenance which is planned, ongoing or finished on the Bavarian street network. The dataset for 2019 contains close to 650.000 data-points which each describe the temporal and spatial extend, road name and number of closed lanes of a roadwork fragment. This fragmentation of events makes it hard to statically analyze this dataset since each roadwork is spitted into any number of fragments which are only linked by a roadwork identifier. Therefore the analysis of the dataset in this section is relatively basic. The import processing works similarly to the BAYSIS data in section 4.2 and produces 282.839 roadwork events in the database after the aggregation of fragments.

With 4500 long term and more than 40.000 short term building sites on German highways per year (LAPID 2018; Stmi 2020) road construction makes up for the majority of traffic obstructions during the summer months. Throughout the colder months, in which a majority of construction projects are not possible, snow clearings or long-term constructions are the main obstacles. This also means that the number and type of roadworks varies during the course of a year (Stmi 2020). The monthly distribution of roadworks in the year 2019 in figure Figure 4.4 supports this state-

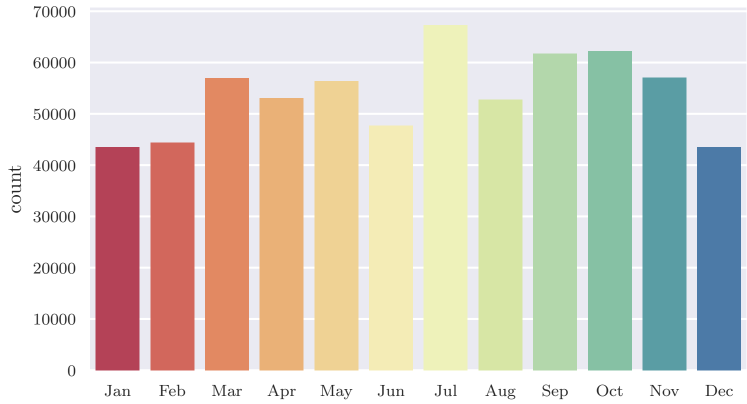


Figure 4.4: Monthly distribution of roadwork fraction counts

ment, since the winter months of January, February and December tend to have less roadwork that others. In July most of the roadwork is done. Similar to the BAYSIS data the road *A3* and A9 have the highest numbers of roadworks (shown in figure Figure 4.5).

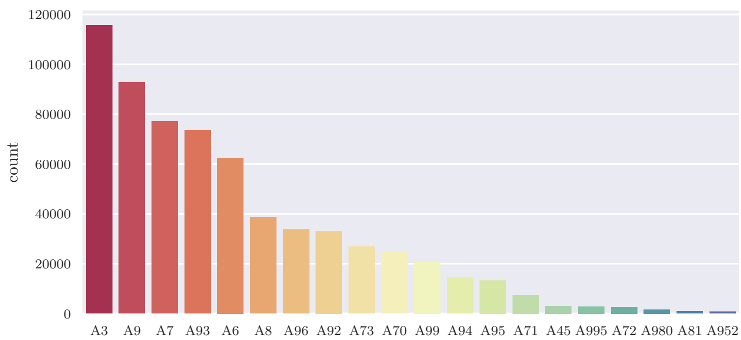


Figure 4.5: Distribution of roadwork fraction counts, by road

AnzGesperrtFs refers to the number of closed lanes for the time of the incident. The distribution shows that two states of zero and one block lane hold nearly 100 % of all samples. Since the variable can be ordered by the number of lanes, it is of an ordinal type.

AnzGesperrtFs	Count	[%]
0	215977	33.4
1	430022	66.5
2	140	< 0.1
3	81	< 0.1

Table 4.11: Descriptive of **AnzGesperrtFs**

Einzug describes the shift of the road way due to physical changes measured in the number of lanes. It ranges from one to five lanes, where one, two and five are equally frequent. The variable is limited, can be sorted and therefore is of an ordinal type.

Einzug	Count	[%]
1	230192	35.6
2	201289	31.1
3	728	0.1
4	1	< 0.1
5	214543	33.2

Table 4.12: Descriptive of **Einzug**

Since **Length** and **Duration** are not part of the original dataset, they are calculated from the dataset parameters *VonKilometer* / *BisKilometer* and *Von* / *Bis* respectively. They are measurements and therefore are of an interval type.

5 Methodology of processing

The previous chapters provided an introduction into the systems, data and methods which will be used to determine if jam characteristics detected in FCD and incident characteristics from accidents (BAYSIS) and roadworks (ArbIS) are statistically related to each other. The research questions to be answered can be termed as:

Do congestion- and incident-characteristics correlate?

and

Are congestion- and incident-characteristics predictable?

The methodology to answer this research questions will be elaborated in this chapter, starting with the detection of jams in the FCD. This also contains the generation of congestion characteristics and collection of adjacent incidents. FCD is a continuous time series of datapoint which represents the mean absolute and relative speed of the street section at each 3-minute interval on a road section. In section 4.1 it was determined, that trough a manual visual analysis jams can be easily identified. This manual identification will be automated because of the amount of data representing a complete year and all Bavarian highways. The gathered tuples of congestion and incident are then processed and exported into a unified data-structure.

The evaluation tool of the CONGSTATS service which is the project this thesis was inspired of, was developed for this purpose and will be expanded with required features. Afterwards the stored dataset of congestion and incidents events will be analyzed for correlations and other statistic indicators.

5.1 Detection Algorithm

The first step is the detection of congestion events. In section 2.2 a congestion is defined as a dense, temporal and spatial accumulation of jammed cells, also describable as a cluster of jammed cells. Therefore a clustering algorithm would be suitable to identify congestion events.

A shaping algorithm is needed for the classification of the congestion events into different types by their spatial and temporal extends. It is supposed to convert the accumulation of cells into a simple describable shape which can be put into groups.

5.1.1 Clustering of Floating-Car-Data

The term clustering is a short form of a data mining technique also called numerical taxonomy or cluster analysis with the goal of finding data structures or associations. For this purpose a multitude of algorithms where developed over time, varying in their strategies, methods and performance (Busch 2005). For example k-means or k-metoid (point distance), affinity propagation (graph distance), mean-shift (point distance), DBSCAN (nearest point distance), gaussian mixtures (mahalanobis distance to centers) or spectral clustering (graph distance), which can be sorted into the categories of partition-based, hierarchical-based and density-based clustering (Chauhan 2020; Yildirim 2020)

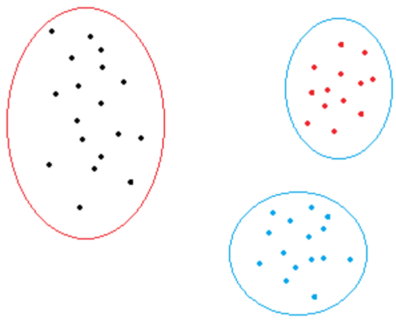


Figure 5.1: Example clustered by k-means algorithm (Yildirim 2020)

To illustrate which problems can occur when using cluster algorithms and to define the features which makes an algorithm suitable, the k-means is used as base comparison. The fig. 5.1 shows three differently colored groups of points which are grouped into three clusters by the k-means algorithm, represented by the colored circles around the groups. It demonstrates the general principle of clustering and was done by the common k-means algorithm with the a priori parameter of three clusters.

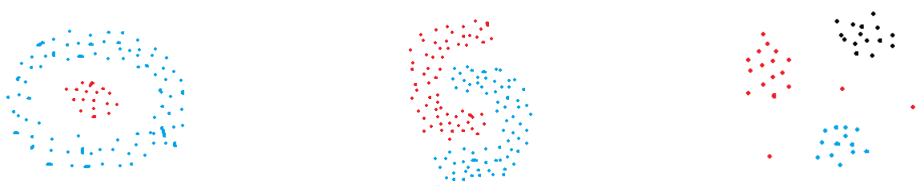


Figure 5.2: Example clustered by density-based algorithm (Yildirim 2020)

This algorithm produces representable results as long as the groups don't overlay, intersect or have arbitrary shapes like in fig. 5.2 which was clustered by the DBSCAN algorithm. If this is the case, the k-mean may cluster loosely related points together which actually are more strongly related to other points, because it considers every point as a possible neighbor and other algorithms are better suited. Since jams in FCD can overlay and appear in any shape, a suitable algorithm is needed to be able to handle such data.

Entering density-based methods which are better suited to identify distinctive, arbitrary clusters, by looking for a contiguous region of high point density, separated from others by contiguous regions of low point density (Chauhan 2020).

Density Clustering Algorithm

The DBSCAN algorithm meaning **d**ensity-**b**ased **s**patial **c**lustering of **a**pplications with **n**oise, is able to find arbitrary shaped cluster and clusters by considering the spatial density, which also represents noise. The basic idea of this algorithm is to form cluster of points, which are close too many other points. For this strategy two threshold parameters are needed. The first being the minimal size of a cluster, referred to as $MinPts$, which defines the minimum number of points necessary to form a cluster. And secondly the maximum distance threshold between points, *eps* (ε), to be considered as neighbors and become part of a cluster. These thresholds classify a data point as a core point, a (directly) density-reachable point or noise. (Chauhan 2020; Prado 2017; Yildirim 2020)

- A **core** point q has at least $MinPts$ points around it within the neighborhood ε, including itself.
- **Directly density-reachable** border points have at least one core point within the neighborhood ε.
- **Density-reachable** border points have at least one core point within the neighborhood ε of a chain of points $p_1, p_2, ...p_n$.
- **Noise** or outliner point are neither core points nor are they density-reachable and therefore have less than $MinPts$ in their neighborhood ε, including themselves.

The general procedure of the algorithm, with the input of n points, neighborhood radius ε and density threshold $MinPts$ is the following (Zhao et al. 2018):

1. Mark all points as *unvisited*

2. Choose point p randomly from all *unvisited* points.

   a. Choose point p randomly from all *unvisited* points and mark p as *visited.*

   b. Count points in the neighborhood ε to check if p is core point. If p is core point, create new cluster C and add all directly density-reachable and *unvisited* points. Otherwise mark p as noise.

   c. Choose point p' randomly from all *unvisited* points of C and mark p as *visited.*

   d. Count points in the neighborhood ε to check if p' is core point. If p is core point add all directly density-reachable points which do not already belong to a cluster to C. Otherwise mark p as noise.

e. Repeat step **c** and **d** until there are no *unvisited* points left in C.

3. Repeat step **2** until all points are *visited*

Artificial Distance Measuring

The algorithm is based on the density of points. This density representation is achieved via checking the neighborhood ε against the calculation of distance from point A to B. Typically the Euclidean distance, which is based on the pythagorean theorem is used (see equation 5.1 (Erhard 2020)).

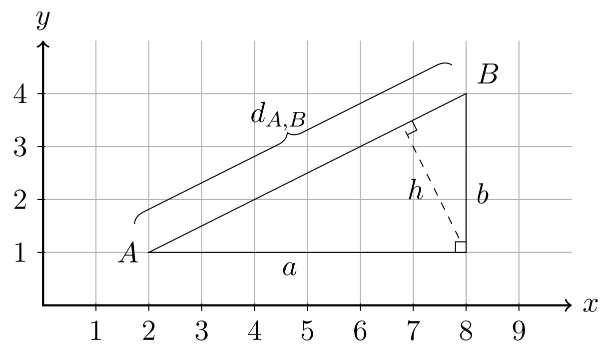


Figure 5.3: Euclidian distance in 2-dimensional euclidian space

$$d_{A,B} = \sqrt{a^2 + b^2} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} \tag{5.1}$$

As shown above the Euclidean distance calculation assumes a Euclidean space and therefore the equal scaling of both axis, which are not given with provided FCD. The 2D-space of the FCD is defined by a spatial and a temporal dimension which are scaled differently as already mentioned in section 4.1. The vertical axis shows the temporal dimension is scaled regularly in 3 minute intervals. The horizontal axis is the spatial extend, scaled in one cell per step, representing one road link (road links are rather small subsections of roads, defining the course), which vary heavily in their length. This can be visualized like in the following fig. 5.4:

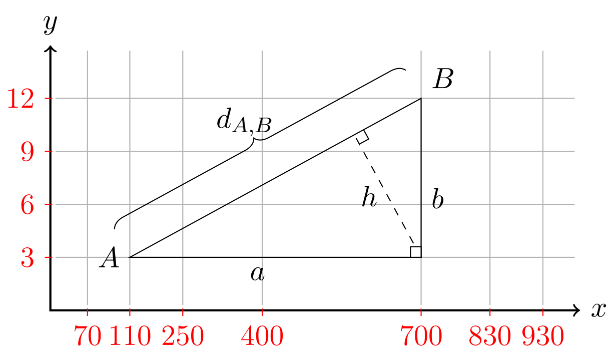


Figure 5.4: Euclidian distance in 2-dimensional non-Euclidian space

This makes a direct application of the Euclidian distance invalid to calculate the distance from point A to point B. But then how can distances be represented in this 2D-space with different units? As a solution the 2D representation of time and space can be expanded into a 3D-representation of time, space and speed (absolute mean traveling speed is part of FCD). Because speed $[km/h]$ consist of both time $[min]$ and space $[m]$ it allows for the introduction of the common unit travel time which can be comprised from these three parameters traveling speed, time duration and link length. At this point it has to be noted that the implemented definition does not represent actual travel time but server the need of being common unit and rough representation. To separate both terms the term *artificial travel time* will be used instead of *travel time.*

Assuming an 3D-space of *timesteps* $\equiv x = [x_1, ..., x_n]$, *space* $\equiv y = [y_1, ..., y_n]$ and *speed* $\equiv z = [z_1, ..., z_n]$ the mathematical definition of calculation for the distance from point A to point B is as follows.

Time dimension : Traveling just in the time dimension is not actually possible, hence the term artificial travel time. But assuming that both points A and B are at the same **space** step, the distance is calculated by:

$$t_{x,A,B}^{artificial} = (x_B - x_A) \cdot (x_{interval} + 1) \tag{5.2}$$

$x_A|x_B$ is the x index of point $A|B$

$x_{interval}$ is the time interval or step duration

The listing 5.1 shows the evaluation tool implementation of the the vertical distance calculation.

Listing 5.1: Implementation of *vertical distance calculation*

```java
/**
 * Computes the vertical distance from A to B in [s]
 *
 * @param a The indexes of from where, with [0] = stepIdx and [1] = linkIdx
 * @param b The indexes of to where, with [0] = stepIdx and [1] = linkIdx
 * @return The travel time in seconds
 */
public double computeVertical(final int[] a, final int[] b) {
    // make sure that cells are on the same link
    assert a[1] == b[1];
    // calculate number of steps between cells
    final int deltaSteps = (Math.abs(b[0] - a[0]));
    // the vertical travel time in [s] is sum of ( time steps [idx] * step duration [min] * 60 [s] ) and
        therefore is inclusive
    return deltaSteps * stepDuration * 60;
}
```

Space dimension : Traveling just through space is also not actually possible, but only in the artificial travel time. Under the assumption that point A and point B are in the **time** step, the distance is calculated by:

$$t^{artificial}_{y,A,B} = (\frac{y^{length}_A}{2} \cdot z^{speed}_A) + (\frac{y^{length}_B}{2} \cdot z^{speed}_B) + \sum_{i}^{y_A+1,...,y_B-1} (y^{length}_i \cdot z^{speed}_i) \tag{5.3}$$

$y_A|y_B$ is the y index of point $A|B$

y_{length} is the length of the link

z_{speed} is the speed in the cell

The listing 5.2 shows the evaluation tool implementation of the the vertical distance calculation.

Listing 5.2: Implementation of *horizontal distance calculation*

```
/**
* Computes the horizontal distance from A to B in [s]
*
* @param a_idx The indexes of from where, with [0] = stepIdx and [1] = linkIdx
* @param b_idx The indexes of to where, with [0] = stepIdx and [1] = linkIdx
* @return The travel time in seconds
*/
 public double computeHorizontal(final int[] a_idx, final int[] b_idx) {
   // make sure that cells are in the same time step
   assert a_idx[0] == b_idx[0];
   final int stepIdx = a_idx[0];
   // space travel time in [s] is sum of ( link length [m] / ( the speed [km/h] * 1000 / 3600 ) [m/s] )
   double spaceTime = 0;
   int lowerLinkIdx = Math.min(a_idx[1], b_idx[1]);
   int upperLinkIdx = Math.max(a_idx[1], b_idx[1]);
   for (int i = lowerLinkIdx + 1; i < upperLinkIdx; i++) {
     double travelSpeed = speedMatrix[stepIdx][i] * 1000D / 3600;
     spaceTime = spaceTime + (linkLengths[i] / travelSpeed);
   }
   // add half travel time from start and end cell (middle to middle or inclusive)
   double travelSpeedA = speedMatrix[stepIdx][lowerLinkIdx] * 1000D / 3600;
   spaceTime = spaceTime + (linkLengths[lowerLinkIdx] / travelSpeedA / 2);
   double travelSpeedB = speedMatrix[b_idx[0]][upperLinkIdx] * 1000D / 3600;
   spaceTime = spaceTime + (linkLengths[upperLinkIdx] / travelSpeedB / 2);
   return spaceTime;
 }
```

Diagonal time/space dimension : The diagonal distance or time/space distance is based on the concept of the Euclidian distance. Although the conversion to the artificial travel time fixes the disparity of the axis units, a calibration for weighing the time and space axis is still necessary. This can be achieved by using the aimed gap thresholds to form a calibrator value which scales the axis appropriately to be of equal scaling. This makes it possible to use the artificial travel time a distance parameters, for horizontal, vertical and diagonal movements.

$$t^{artificial}_{x,y,A,B} = \sqrt{(t^{art}_{x,A,B})^2 + (t^{art}_{y,A,B} \cdot c)^2} \tag{5.4}$$

$$v^{mean}_{A,B} = \sum_{ij}^{xy_A+1,...,xy_B-1} z^{speed}_{ij} \tag{5.5}$$

$$c = \frac{t_{min,gap} \cdot v_{freeflow}}{l_{min,gap}} \tag{5.6}$$

$xy_A|xy_B$ is the xy index of point $A|B$

z_{speed} is the speed in the cell

v^{mean} is mean speed in the area between A and B

c is the time-space calibrator

$t_{min,gap}$ is the the aimed time gap

$l_{min,gap}$ is the the aimed space gap

$v_{freeflow}$ is the assumed free flowing speed (typically 130 [km/h])

The listing 5.3 shows the evaluation tool implementation of the the vertical distance calculation.

Listing 5.3: Implementation of *diagonal distance calculation*

```java
/**
* Computes the distance between two n-dimensional vectors.
* <p>
* The vectors are not required to have the same dimension, but should represent the
* [0] time step and [1] link step
*
* @param a The first vector with [0] = stepIdx and [1] = linkIdx
* @param b The second vector with [0] = stepIdx and [1] = linkIdx
* @return The distance between the two vectors
* @throws DimensionMismatchException If the array lengths differ.
*/
 public double computeDistance(int[] a, int[] b) throws DimensionMismatchException {
   if (a.length != b.length) {
     throw new DimensionMismatchException(a.length, b.length);
   }
   if (a[0] == b[0] && a[1] == b[1]) {
     // same link and same step -> travel time from A to B is zero
     return 0;
   } else if (a[1] == b[1]) {
     // same link -> travel time from A to B can be calculated vertically
     return computeVertical(a, b);
   } else if (a[0] == b[0]) {
     // same step -> travel time from A to B can be calculated horizontally
     final double minimalVerticalTravelTime = stepDuration * 60 * 0.5;
     return Math.sqrt(Math.pow(minimalVerticalTravelTime, 2) + Math.pow(computeHorizontal(a, b) *
          horizontalCalibrator, 2));
   } else {
     // compute diagonal with pythagorus and mean of speeds
     final int sigT = a[0] > b[0] ? -1 : 1;
     final int sigX = a[1] > b[1] ? -1 : 1;
     final int deltaSteps = sigT * (Math.abs(b[0] - a[0]) - 1);
     final int deltaLinks = sigX * (Math.abs(b[1] - a[1]) - 1);
     // collect absolute speed in rectangle from cell A and cell B
     final int[] absoluteSpeeds = new int[(Math.abs(deltaSteps) + 2) * (Math.abs(deltaLinks) + 2)];
     int absoluteSpeedsCounter = 0;
     for (int i = 1; i <= Math.abs(deltaSteps) + 2; i++) {
       for (int j = 1; j <= Math.abs(deltaLinks) + 2; j++) {
         absoluteSpeeds[absoluteSpeedsCounter] = speedMatrix[a[0] + sigT * i - sigT][a[1] + sigX * j - sigX];
         absoluteSpeedsCounter++;
       }
     }
     // compute mean travel time over cell rectangle
     int speedSum = 0;
     for (int i = 0; i < absoluteSpeedsCounter; i++) {
       speedSum += absoluteSpeeds[i];
     }
```

```java
        final double meanSpeed = (double) speedSum / absoluteSpeedsCounter;
        // compute the total space between the points in [m] by adding up all link lengths
        final int deltaSpace;
        if (deltaLinks == 0) {
          deltaSpace = 0; // if the cell A and B are right next to each other -> no spacial distance
        } else if (sigX < 0) {
          int tempDeltaSpace = 0;
          tempDeltaSpace += linkLengths[a[1]] / 2; // add beginning cell length half
          for (int i = b[1] + 1; i < a[1]; i++) {
            tempDeltaSpace += linkLengths[i];
          }
          tempDeltaSpace += linkLengths[b[1]] / 2; // add ending cell length half
          deltaSpace = tempDeltaSpace;
        } else {
          int tempDeltaSpace = 0;
          tempDeltaSpace += linkLengths[a[1]] / 2; // add beginning cell length half
          for (int i = a[1] + 1; i < b[1]; i++) {
            tempDeltaSpace += linkLengths[i];
          }
          tempDeltaSpace += linkLengths[b[1]] / 2; // add ending cell length half
          deltaSpace = tempDeltaSpace;
        }
        // compute the total space time in [s] from the link lengths and absolute speeds
        final int deltaSpaceTime;
        deltaSpaceTime = (int) Math.round(deltaSpace / (meanSpeed * 1000 / 3600));
        // compute the total time between the points in [s], can be done vertically
        final double deltaTime = computeVertical(new int[]{a[0], a[1]}, new int[]{b[0], a[1]});
        // compute and return the mean travel time from A to B
        return Math.sqrt(Math.pow(deltaTime, 2) + Math.pow(deltaSpaceTime * horizontalCalibrator, 2));
      }
    }
```

Performance Tuning

A considerable performance issue of clustering algorithms is the runtime complexity when processing larger amounts of data. It is defined with the Big O notation, describing the mathematical runtime complexity of an algorithm. Leaving out the runtime implication of the distance measurer, the complexity of DBSCAN clustering algorithm can be as low as $O(nlog_n)$. This best case scenario can be achieved by using indexing systems to store the clustering data in a space representation like a 2D-Tree. This reduces the number of points to check for neighborhood, from all points in a worst case scenario (equivalent to a complexity of $O(n^2)$), to just adjacent points (Chauhan 2020). For initial testing a proprietary *kd*-Tree was implemented, which stores data points as leafs in a tree, where the nodes divide the space consecutively in the x and y dimension (Dalitz et al. 2009; Hucker 2020). This improved the runtime performance, but also added complexity, which endorsed the use of a natively implemented data structure, like the TreeMap. The TreeMap strictly speaking is not a 2D-Tree, but has an average complexity of $O(nlog_n)$ and be used as a 2D-Tree by filtering with two parameters (see source code in repository *code/congestion/clustering/pool*) (Baeldung 2020a,b).

To further accelerate the algorithm it is implemented with support of parallel computation or threading which allows the executing Java VM to used multiple CPU cores and run multiple processes in parallel (see source code in repository.

Calibration

Parameter calibration and estimation is a vital task when implementing algorithms. The DB-SCAN uses the neighborhood ε and $minPoints$ parameters as adjustments. If ε is too small, part of the data will not be clustered, since the distances to many points is below the threshold. These points are therefore considered as outliner/noise and reduce the actual size of the cluster or make the cluster neglect able because $minPoints$ will not be reached to create a dense region. On the other side, if the value is chosen too high, a high number of points will be considered as one cluster, when they should be multiple separate clusters. The $minPoints$ threshold should generally satisfy $minPoints > D + 1$ and should be high enough for our implementation to neglect small and arbitrary jams. (Prado 2017). For the implemented variation of distance measuring the aimed time gap $t_{min,gap}$ and aimed space gap $l_{min,gap}$ threshold also need to be set. This is necessary to scale the axis so that they represent the aimed thresholds through neighborhood ε. The following values are the result of iterative testing to find the most representable cluster consolidation results.

- Aimed spatial gap : $l_{min,gap} = 5000\,[m]$
- Aimed temporal gap : $t_{min,gap} = 6\,[min]$
- Virtual travel-time gap $\varepsilon = t_{min,gap}^{artificial} = 360\,[s]$

5.1.2 Pre and Post Data Revision

Datasets are rarely flawless and as mentioned in section 4.1, the provided FCD dataset has some defects. To reduce these defects before the clustering, static speed blocks are removed as pre-processing measure. When the speed is consistent in a continuous time and space extent, it can be assumed that the data block is flawed because of the implausible consistent speeds.

As post processing, clusters which are too short in duration and length are removed. It was assumed that below a threshold a cluster can not be considered as a jam and should be neglected. The following values for the minimal duration and length of a congestion where used.

- Minimum spatial length of an congestion event : $l_{min} = 1000\,[m]$
- Minimum temporal duration of an congestion event : $t_{min,gap} = 9\,[min]$

5.1.3 Shaping

For a shape representation of the congestion the geometric method of convex hull was implemented. This shape was initially intended to be the base for the classification of congestion event into different type of jams alongside the paper *Automated Classification of Different Congestion Types* (Kessler et al. 2020). Unfortunately this classification processing was not finished in time for this thesis, but the implementation found use in the visual representation of jams and characteristics calculation.

5.2 Matching Algorithm

The matching process for finding adjacent incidents around jams is rather simple. When iterating over all jams, the incidents located on the same road and on the same day are evaluated for the temporal and spatial distance to the outer line of the the congestion. When the distance falls in the range of the threshold they are considered as adjacent. The following values where used as thresholds.

- The spatial distance an adjacent incident : $l_{min,dist} = 2000\,[m]$
- The temporal distance an adjacent incident : $t_{min,dist} = 25\,[min]$

The final implementation of the detection and clustering algorithm of the evaluation tool can be reviewed in *code/congestion/* of the repository which is linked in the introduction.

5.3 Data Processing

As a result of the previous detection and matching algorithms a list of congestion objects with spatial and timely adjacent incidents objects (accidents and roadworks) was created. For the statistical analysis these congestion and incident matches will be expanded with additionally features and exported into a local data format.

For the analysis the length and duration of the congestions is of interest and therefore needs to be defined and calculated. When defined by the boundary rectangle the two measurements can be heavily biased. It is also a very rough representation of the extends and is therefore considered to be the maximum length and maximum duration. To have another representation of the time and space extends, an average duration and length is calculated. This is done by iteration over the time- or link step and calculating the mean of the jam length or duration respectively.

For the analysis of social impact the congestion object is expanded with a time loss estimation, differentiated for passenger cars and heavy goods vehicles (HGV). This part was not implemented by the writer, but from the mentor Stefan Gürtler (S&W) and was used as implemented. Since it is used in the analysis, the calculation will be explained in the following. First the headway with the cell speed is calculated with the assumption that drivers follow the two second rule, see eq. (5.7).

$$l_{headway} = 2 \cdot \frac{v_{cell}}{3600} \tag{5.7}$$

Then the occupied space of a hundred vehicles can then be comprised by the headway and the vehicle lengths as in eq. (5.8).

$$l_{100,vehicles} = n_{car} \cdot (l_{headway} \cdot l_{car}) + n_{hgv} \cdot (l_{headway} \cdot l_{hgv}) \tag{5.8}$$

$$n_{car} = (1 - r) \cdot 100 \tag{5.9}$$

$$n_{hgv} = r \cdot 100 \tag{5.10}$$

The density of vehicles for the cell speed can be calculated as in eq. (5.11).

$$d = \frac{1}{((1 - r) \cdot (l_{headway} \cdot l_{car}) + r \cdot (l_{headway} \cdot l_{hgv})} \tag{5.11}$$

The appropriate vehicle count can then be deviated by eqs. (5.12) to (5.14)

$$n_{vehicles} = l_{cell} \cdot d \tag{5.12}$$

$$n_{car} = (1 - r) \cdot n_{vehicles} \tag{5.13}$$

$$n_{hgv} = r \cdot n_{vehicles} \tag{5.14}$$

To calculate the total hours lost due to the jam, for either passenger car or heavy goods vehicle, eq. (5.15) is applied with the values calculated before.

$$t_{loss,car/hgv} = n_{car/hgv} \cdot n_{lanes} \cdot l_{cell} \cdot (v_{free} - v_{cell}) \tag{5.15}$$

v_{cell} is the mean vehicle speed in [km/h]

v_{free} is the assumed free flowing speed in [km/h]

d is the density of vehicles in the cell [veh/km]

r is the ratio of heavy goods vehicle to passenger cars

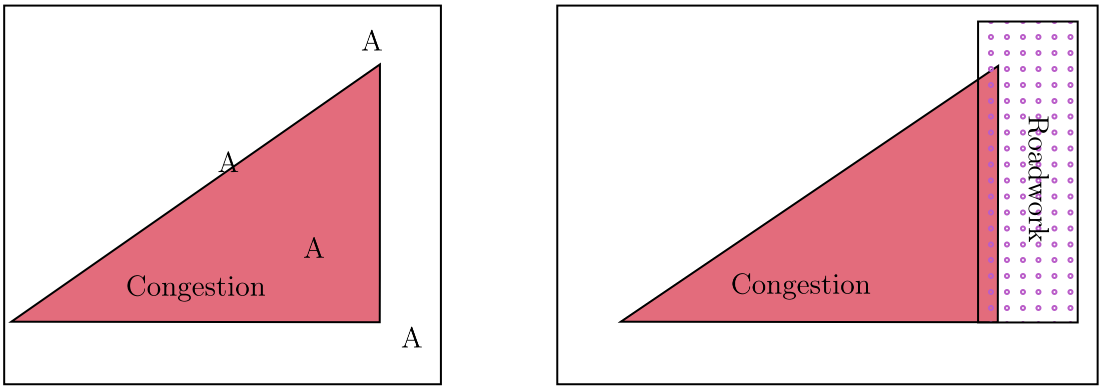
An analysis based on all matches could be heavily biased. For a more specialized analysis the congestion - incident matches should be categorizable into different relation types like initiation or effect. For accidents the three categories to be evaluated separately can be described as *Initiators*, *Effectors* and *Followers*. For roadworks the same detailed categorization is not that viable, but a single selection of *Initiators* is.

Jam Initiators are matches where the accident happened before or immediately at the beginning of a congestion. These could be the initiator or cause for the congestion and are subjectively the most interesting kinds for analysis. In terms of roadwork these are matches where the roadwork is located after or during the congestion and the roadwork could be considered as cause.

Jam Effectors are matches where the accident happened during a congestion or overlaps with the congestion and probably have the jam as a cause, like rear-end accidents.

Jam Follower are matches where the accident happened after the congestion.

This classification can be visually represented like in fig. 5.5, where the red triangle symbolizes a congestion in the FCD space. The three classifications *Initiator*, *Effector* and *Follower* are symbolized with the different hatchings of yellow stripes, green squares and red rhombus. To



(a) Accidents **(b)** Roadworks

Figure 5.5: Visual representation of the accident and roadwork classifications *Initiator* (yellow stripes), *Effector* (green squares) and *Follower* (red rhombus)

be able to categorize after the runtime intensive detection and matching processing, which allows easier testing and calibration, four additional congestion parameters are introduced into the congestion object. They describe the relative location of the accident to the congestion event. The temporal reference of the accident to the congestion is described by the parameter *temporalGlobalLocation* (shown in table 5.1), based on the temporal distance. The spatial reference

1	is before
2	is overlapping before
3	is during
4	is overlapping after
5	is after

Table 5.1: Encoding and description of temporal global location reference

of the incident to the congestion is described by the parameter *spatialGlobalLocation* (shown in table 5.2), based on the spatial distance. In case that the incident happened temporal during the

1	is before
2	is during or overlapping
3	is after

Table 5.2: Encoding and description of spatial global location reference

congestion. The (*temporalInternalLocation* (shown in table 5.3) parameter is set. It describes percentile thresholds where the incident is located during the congestion. In case that the inci-

1	10 % to Beginning
2	10 % - 30 % to Beginning
3	30 % - 70 % (Middle)
4	30 % - 10 % to Ending
5	10 % to Ending

Table 5.3: Encoding and description of temporal internal location reference

dent happened spatial during the congestion. The *spatialInternalLocation* (shown in table 5.4) parameter is set. It describes is percentile thresholds where the incident is located during the congestion.

1	10 % to Beginning
2	10 % - 30 % to Beginning
3	30 % - 70 % (Middle)
4	30 % - 10 % to Ending
5	10 % to Ending

Table 5.4: Encoding and description of spatial internal location reference

After the processing and expansion a congestion object contains the following attributes.

Name	Unit	Description
TMax	*min*	Temporal maximal extend, based on boundary rectangle
TAvg	*min*	Temporal maximal extend, based on boundary convex hull
SMax	*m*	Spatial maximal extend, based on boundary rectangle
SAvg	*m*	Spatial maximal extend, based on boundary convex hull
TDist	*min*	Temporal minimal distance
SDist	*m*	Spatial minimal distance
Cov	%	Coverage of jammed area of the boundary rectangle
TLCar	*h*	Total time loss of passenger cars
TLHGV	*h*	Total time loss of heavy goods vehicles

Table 5.5: Variables, units and descriptions of congestion object

The incident objects for accident and roadworks contain the variables described in the section 4.2 and section 4.2.

5.4 Correlation Processing

The correlation processing is written in Python and R with the help of some common data analysis frameworks like Pandas, NumPy or Psych. The code base was inspired by the analysis tool from Potvin (Potvin 2020) and can be review in the repository linked in the introduction. This section will explain the process of analyzing the two datasets from the evaluation processing (roadwork-congestion and accident-congestion). The first step in any data analysis is the data import and preparation. After these initial steps the tool calculates the correlations and significances for all variable combinations. These step are applied to each of the two datasets of congestion - accident and congestion - roadwork as well as their respective subsets of *Initiator*, *Effector* and *Follower* (see *./code/main_ ... .py* files).

Data Revision

The data processing in the evaluation tool already cleaned the ingress data from most defects, however it does not remove not defective but flawed data samples. During the analysis it became clear that the congestion detection still produces congestions with abnormal and non plausible extends. As a preprocessing of the correlation analysis, jams with a length over 50 km are removed, to remove samples with abnormal length like 200 km and beyond. Also empty values and -1 values are considered as not available (*NaN*) and are removed from the sample set.

Data Encoding

The statistical methods to be used and defined in section 3.2 can only be applied on numerical data (not to be confounded with the variable type nominal from section 3.1). Therefore the variables which contain characters or other non-numerical data need to be encoded. This is done by assigning an increasing integer to a non-numerical variable referring to a group. This increasing integer and the corresponding group is then saved in a code - group dictionary to be able the refer back to the original group.

Correlation Calculation

The correlation calculation is the implementation of the statistical methods which are defined in section 3.2. The Python function *compute_ correlations* (see *./code/func_ correlation.py*) takes the encoded dataset and computes the effect size of each appropriate correlation coefficient for each variable relation, from which an initial group of correlated characteristics can be identified.

Significancy Evaluation

As described in section 3.5.2 the correlated relations need to be further tested for significance. This is done via the post hoc test and is implemented in the R scripts *baysis_ ... .R* and *arbis_ ... .R* (see *./code*). The script takes the correlated characteristics and tests the relation for significance and significant differences as defined in section 3.6. The results of the post hoc test provide a definitive answer about the association and predictability of the relation.

6 Analysis of processing results

In this chapter the result of the data processing will be evaluated. Starting with the evaluation processing, which clusters the FCD, forms congestion events, finds adjacent incident and exports a list of congestion-incident matched. The second section will elaborate on the results of the correlation processing and use the results for a further analysis of relations.

6.1 Evaluation Processing

The results of the clustering and matching algorithm where visually reviewed to verify the performance. Thought iterative adjustments of the input parameters the clustering and matching algorithm where calibrated to a sufficient representation level (see final input parameters in section 5.1 and section 5.2). This sufficient representation is visually presented in fig. 6.1 which

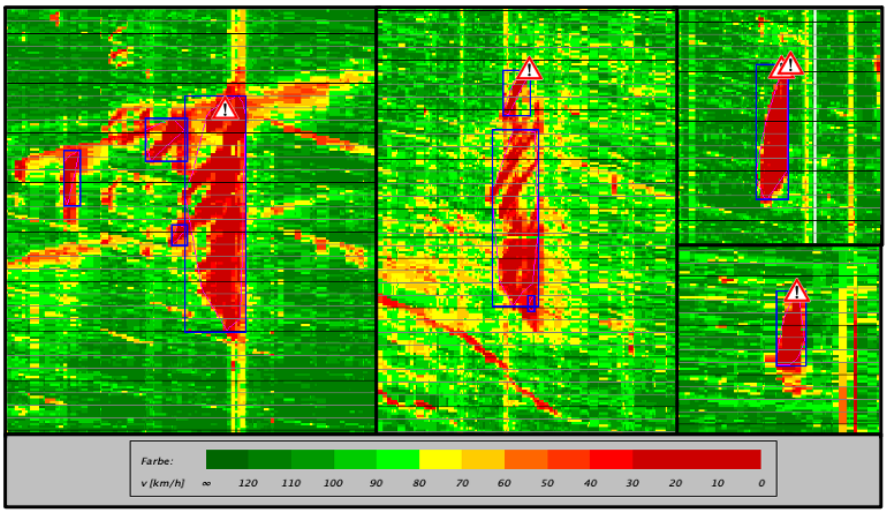


Figure 6.1: Speed matrix plots of FCD data after clustering

show some random examples of the clustering (blue rectangle = bounding rectangle, magenta polygon = convex hull) and incident matching (attention sign = accident). There are however certain problematic scenarios when clusters like fragmentation or oversizing which will be further discussed in the last chapter with other improvements.

6.2 Correlation Processing

The resulting datasets created by the evaluation tool (see section section 5.1 section 5.2 and section 5.3) which is tasked with the detection and clustering of jams and search for adjacent incidents, are then processed by the correlation tool (section 5.4). The correlation tool calculates multiple matrix tables with the correlation effect size, correlation significance and used correction coefficient for all variable combinations. From these tables and interpretation guidelines defined in section 3.2 for each coefficient type, we can deviated the strength of correlation and significance (see section 3.5.2) for each variable combination. To recap section 3.2, table 6.1 shows the guidelines for a weak, moderate and strong correlation effect size of the coefficients r,η,r_{pq},τ and V. The significance is evaluated by an α-level of .05. In the case of correlated and significant

Coefficient	Weak	Moderate	Strong
r	.30	.50	.80
η	$<$ .06	.06	.14
r_{pq}	$<$ .30	.30	.50
τ	$<$ .30	.30	.50
V	$<$ .30	.30	.40

Table 6.1: Correlation effect size interpretation for coefficient r, η, r_{pq}, τ and V

variables, it still needs to be determined what the found correlation predicates. This is done via the Post Hoc test, defined in section 3.6, which tests for for significance differences between the groups via the pairwise Wilcoxon T-test. The rest of this chapter is dedicated to elaborate on this tedious process of testing all groups for significance differences. This involves a enormous number of tables which need to be evaluated and involves repetitions, but is necessary to cover all assumptions and interpretations, referenced later on. For a summary of the significant differences and their interpretations, please forward to chapter 7.

6.2.1 Congestion - Accidents in general

The correlation matrix table for the complete congestion - accident matched dataset (see table A.17) is visual presented in fig. 6.2 showing the the correlation of each variable combination. When visual analyzing fig. 6.2 and checking the guidelines for a strong correlation in reference to the applied coefficient (identifiable with table A.19) we get a list of strongly correlated variable combinations (see table 6.2). Since the focus of the thesis are the correlations between accidents and jams, these are only collected from the bottom-left corner of the matrix, where the congestion and accidents variables intersect. Correlations of the kind congestion - congestion or accident - accident are not considered. Next we need to verify that the correlation is significant and what the correlation predicates. Therefore each correlation will be evaluated with the Post Hoc test, defined in section 3.6. In the following sections, the correlated relations of the variables in table 6.2 are analyzed and an initial interpretation of each significant correlation is introduced. Groups with an insufficient sample size (see section 3.5.1 are neglected and not shown to reduce

Category	Strong
Str	TMax, TAvg, SMax, SAvg, TDist, SDist, Cov
Kat	TMax, TAvg, SAvg, TDist
Typ	TDist, Cov
UArt1	SAvg, TDist, Cov
AUrs1	SAvg, TDist, SDist, Cov, TLHGV
AufHi	TMax, TAvg, TDist, Cov
Lich1	Cov
Lich2	Cov
Zust1	Cov
WoTag	Cov
Month	Cov

Table 6.2: List of incident variables and their strong correlated congestion variable from the congestion - accident matched data

the overall table size. The descriptive tables, showing the count (n), mean ($\bar{x}$), standard deviation (σ), median ($\tilde{x}$), min, max and range (Δ) therefore only contain groups with significant sample sizes.

Street

This section analyzes the correlated relations of the accident variable *Str*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The Kruskal-Wallis tests of *Str* - *TDist* and *Str* - *SDist* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Str* for these relations are present.

The Kruskal-Wallis test of the relation *Str* - *TMax* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TMax* between the groups of *Str*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.3. It shows

	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A9	0.01	1.00														
A70	0.03	1.00	1.00													
A96	0.00	1.00	0.27	1.00												
A7	0.00	1.00	1.00	1.00	1.00											
A73	0.00	1.00	0.31	1.00	1.00	1.00										
A92	0.00	1.00	0.16	1.00	1.00	1.00	1.00	0.22								
A94	0.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						

Table 6.3: Pairwise Wilcoxon T-test for *Street* and *Maximal Temporal Extent*, see table A.4 for complete table

that the groups A6, A9, A7, A70, A73, A92, A94 and A96 differ significantly from group A3, but there is no distinctive general trend. The significant descriptives from figs. 6.3a and 6.3b show that the $\bar{x}$ duration of the A3 is 72 min - 154 min higher than the $\bar{x}$ of A6, A7, A9, A70, A73, A92, and A94. Therefore it can be interpreted that accidents on the A3 are associated with significantly longer (temporal) jams than on the A6, A9, A7, A70, A73, A92, and A94.

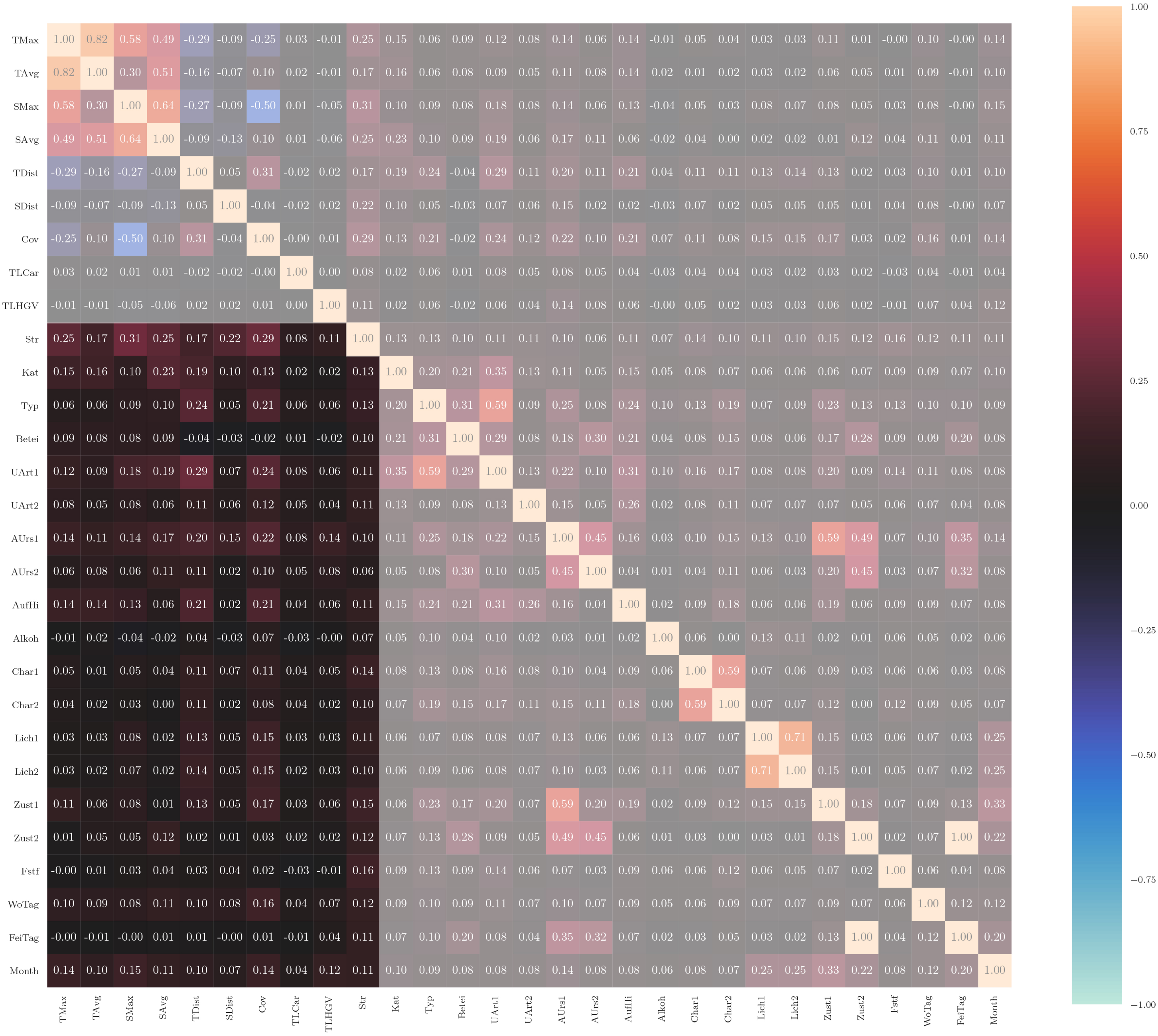
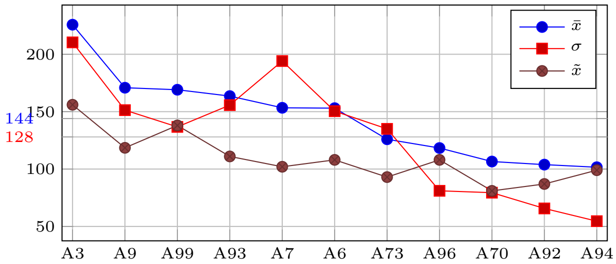


Figure 6.2: Correlation matrix for congestion-accident matched data calculated with V, η, τ, r_{pq}, r

The descriptives also show that the A70, A73, A92, A94 and A96 are have considerable shorter durations, when the A3, A9 and A99 have considerable longer duration compared to the overall $\bar{x}$. The Kruskal-Wallis test of the relation *Str* - *TAvg* produces a p-value of 0.0004, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TAvg* between the groups of *Str*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.4. The table shows

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	559	225.76	210.36	156.0	9	1323	1314
A6	127	153.05	150.42	108.0	12	864	852
A9	466	170.85	151.33	118.5	9	1194	1185
A70	31	106.55	79.42	81.0	24	369	345
A96	155	118.32	81.05	108.0	12	384	372
A7	130	153.37	194.10	102.0	9	1341	1332
A73	129	125.95	135.01	93.0	12	1323	1311
A99	116	169.09	136.72	138.0	15	681	666
A92	66	103.86	65.69	87.0	18	354	336
A93	21	163.57	155.71	111.0	36	588	552
A94	37	101.59	54.60	99.0	15	249	234

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.3: Group descriptives of *Street* and *Maximal Temporal Extent*

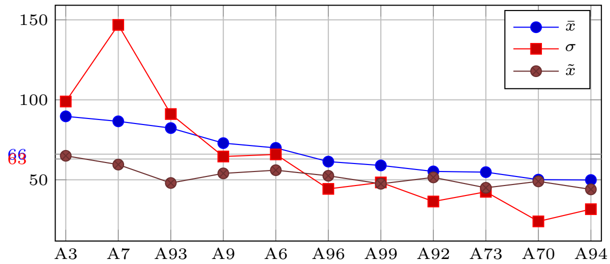
	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A73	0.00	1.00	0.51	1.00	1.00	1.00										
A99	0.02	1.00	1.00	1.00	1.00	1.00	1.00									

Table 6.4: Pairwise Wilcoxon T-test for *Street* and *Average Temporal Extent*, see table A.5 for complete table

that just the groups A73 and A99 differ significantly from group A3, but there is no distinctive general trend. The significant descriptives from figs. 6.4a and 6.4b show that the $\bar{x}$ value of

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	559	89.66	98.94	65.00	4	1260	1256
A6	127	69.94	65.86	56.00	3	376	373
A9	466	72.92	64.55	54.00	4	575	571
A70	31	50.10	23.99	49.00	10	99	89
A96	155	61.37	44.31	52.50	5	247	242
A7	130	86.55	146.82	59.50	6	1326	1320
A73	129	54.78	42.48	45.00	6	274	268
A99	116	58.97	48.35	47.50	4	295	291
A92	66	55.24	36.43	51.50	8	235	227
A93	21	82.33	91.10	48.00	7	343	336
A94	37	49.86	31.63	44.00	14	145	131

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.4: Group descriptives of *Street* and *Average Temporal Extent*

A3 is about 48 min higher than the $\bar{x}$ of A73 and A96. Therefore it can be interpreted that accidents on the A3 are associated with significantly longer (temporal) jams than on the A73 and A96. The descriptives show also that the A3, A7, A9 and A93 are have considerable shorter durations, when the A70, A73, A92, and A94 have considerable longer duration compared to the overall $\bar{x}$. When comparing the $\bar{x}$ values of the maximal and average (temporal) extend

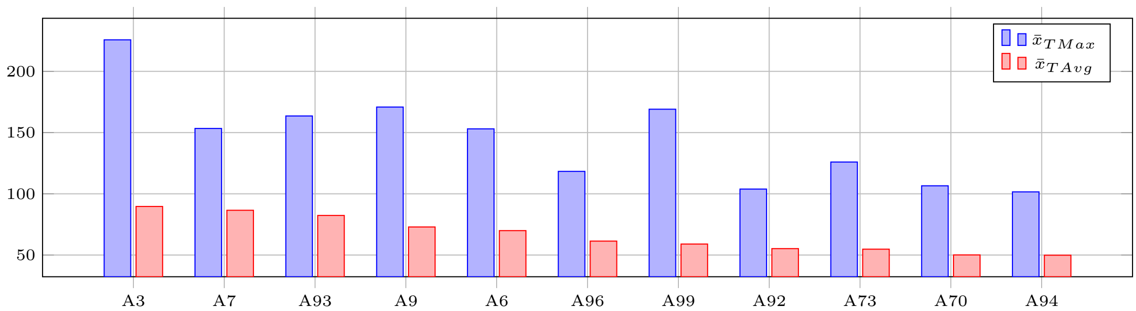


Figure 6.5: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ (*TMax*/*TAvg* by *Street*)

(shown in fig. 6.5) it becomes clear that the average variable has considerable lower values than the maximum variable, which is to be expected. It also shows, that the differences between the groups are mostly similar in both the maximal and average extend, but vary considerably. In can be described that the follow the same tend. This trend is especially ignored by the A99, which probably means that there are more errors in the maximal (temporal) extend variable on the A99.

The Kruskal-Wallis test of the relation *Str* - *SMax* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SMax* between the groups of *Str*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.5. It shows,

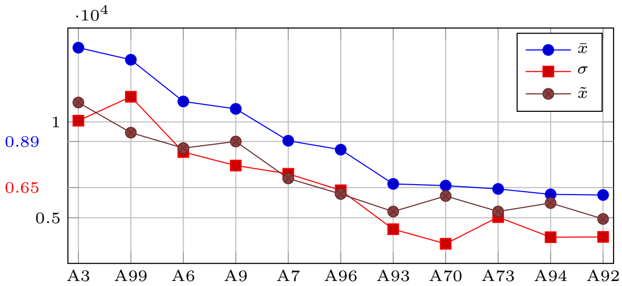
	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A9	0.00	1.00														
A70	0.00	0.83	0.54													
A96	0.00	1.00	0.14	1.00												
A7	0.00	1.00	1.00	1.00	1.00											
A73	0.00	0.00	0.00	1.00	1.00	0.59										
A99	1.00	1.00	1.00	0.80	0.31	1.00	0.00									
A92	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00								
A93	0.03	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A94	0.00	0.11	0.03	1.00	1.00	1.00	1.00	0.09	1.00	1.00						

Table 6.5: Pairwise Wilcoxon T-test for *Street* and *Maximal Spatial Extent*, see table A.6 for complete table

that the groups A7, A9, A70, A73, A92, A94 and A96 differ significantly from group A3. The groups A73 and A92 further differ significantly from group A6 and A9. The group A99 differs significantly from group A73 and group A92 differs from A99, but there is no distinctive general trend. The significant descriptives from figs. 6.6a and 6.6b show that the $\tilde{x}$ of A3 is 3194 m

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	559	13874.96	10064.51	11014.00	1084	46328	45244
A6	127	11067.98	8428.07	8634.00	965	43156	42191
A9	466	10680.48	7724.45	8977.50	832	49765	48933
A70	31	6676.39	3640.24	6136.00	1841	13058	11217
A96	155	8551.75	6431.18	6238.00	971	27965	26994
A7	130	9018.27	7293.14	7051.50	1108	43244	42136
A73	129	6502.88	5033.20	5327.00	1036	33764	32728
A99	116	13244.02	11313.27	9439.50	1280	48278	46998
A92	66	6186.80	4000.10	4936.50	1176	23291	22115
A93	21	6765.00	4403.32	5323.00	1244	16922	15678
A94	37	6220.38	3984.46	5768.00	1167	15550	14383

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.6: Group descriptives of *Street* and *Maximal Spatial Extent*

- 7688 m higher than the $\tilde{x}$ of A7, A9, A70, A73, A92, A94 and A96. They also shows that the groups A6, A9 and A99 have a 5163 m higher $\tilde{x}$ on average than the groups A73 and A92. Therefore it can be interpreted that accidents on the A3, A6, A9 and A99 are associated with significantly longer (spatial) jams than on A7, A73, A92, A94 and A96. The descriptives show also that the A3, A6 A9 and A99 have a considerable longer lengths, when the A70, A73, A92, A93 and A94 have considerable shorter lengths compared to the overall $\bar{x}$. The Kruskal-Wallis test of the relation *Str* - *SAvg* produces a p-value of 0.0027, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *SAvg* between the groups of *Str*. The significant groups can be identified with a pairwise

Wilcoxon T-test, which is shown in table 6.6. It shows, that the groups A70, A73, A92, A93,

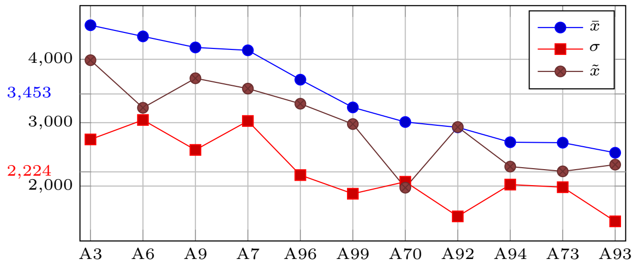
	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A70	0.05	0.83	0.71													
A96	0.05	1.00	1.00	1.00												
A73	0.00	0.00	0.00	1.00	0.00	0.00										
A99	0.00	1.00	0.06	1.00	1.00	1.00	0.51									
A92	0.00	0.61	0.03	1.00	1.00	1.00	1.00	1.00								
A93	0.03	0.46	0.16	1.00	1.00	1.00	1.00	1.00	1.00							
A94	0.00	0.07	0.01	1.00	0.36	0.31	1.00	1.00	1.00	1.00						

Table 6.6: Pairwise Wilcoxon T-test for *Street* and *Average Spatial Extent*, see table A.7 for complete table

A94, A96 and A99 differ significantly from group A3. The groups A73 further differ significantly from group A6, A7, A9 and A96. The A92 also differs significantly from group A9, but there is no distinctive general trend. The significant descriptives from figs. 6.7a and 6.7b show that

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	559	4537.56	2735.62	3986.0	135	17805	17670
A6	127	4361.50	3042.71	3235.0	458	16851	16393
A9	466	4187.15	2569.83	3700.5	393	15132	14739
A70	31	3010.45	2067.63	1974.0	1008	9937	8929
A96	155	3678.39	2172.99	3299.0	387	10182	9795
A7	130	4141.68	3026.58	3537.5	643	16571	15928
A73	129	2683.97	1981.91	2232.0	544	11832	11288
A99	116	3240.97	1878.87	2978.0	583	8426	7843
A92	66	2926.15	1521.59	2931.5	455	8970	8515
A93	21	2525.43	1443.21	2338.0	664	6779	6115
A94	37	2691.68	2023.38	2307.0	358	10393	10035

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.7: Group descriptives of *Street* and *Average Spatial Extent*

the $\bar{x}$ of A3 is 859 m - 2012 m higher than the $\bar{x}$ of A70, A73, A92, A93, A94, A96 and A99. They also shows that the groups A6 and A9 have a 1034 m - 1312 m higher $\bar{x}$ on average than the groups A73 and A92. It can be interpreted that accidents on the A3, A6, A9 are associated with significantly longer (spatial) jams than on A70, A73, A92, A93, A94, A96 and A99. The descriptives show also that the A3, A6, A9 and A7 have a considerable longer lengths, when the A70, A73, A92, A93 and A94 have considerable shorter lengths compared to the overall $\bar{x}$. When comparing the $\bar{x}$ values of the maximal and average (spatial) extend (shown in fig. 6.8) it

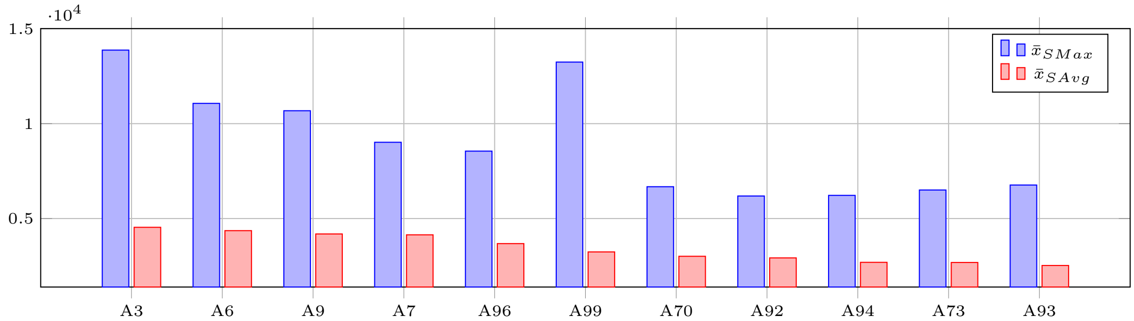


Figure 6.8: Comparison of descriptives $\bar{x}_{SMax}$ and $\bar{x}_{SAvg}$ (*SMax/SAvg* by *Street*)

becomes clear that the average is significantly lower than the maximum, which is to be expected. It also shows, that the difference between the groups are similar in the maximal and average extend. In can be described that the follow the same tend. This trend is only ignored by the

A99, which probably means that there are more errors in the maximal (spatial) extend variable on the A99.

The Kruskal-Wallis test of the relation *Str* - *Cov* produces a p-value of 0.0018, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *Str*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.7. It shows, that the groups A6, A7,

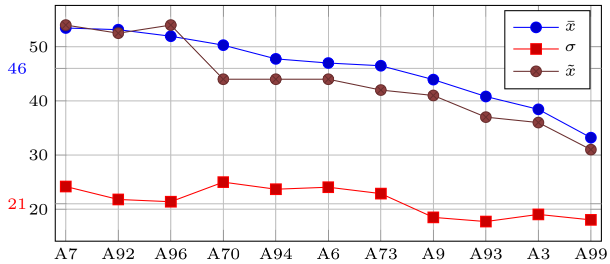
	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A6	0.05															
A9	0.00	1.00														
A96	0.00	1.00	0.00	1.00												
A7	0.00	1.00	0.01	1.00	1.00											
A73	0.04	1.00	1.00	1.00	1.00	1.00										
A99	0.88	0.00	0.00	0.09	0.00	0.00	0.00									
A92	0.00	1.00	0.12	1.00	1.00	1.00	1.00	0.00								

Table 6.7: Pairwise Wilcoxon T-test for *Street* and *Coverage*, see table A.8 for complete table

A9, A73, A92 and A96 differ significantly from group A3. The group A99 differs significantly from group A6, A7, A9, A73 and A96. The groups A7, A92, A96 and A99 differ significantly from group A9. The group A99 also differs from the groups A7, A70, A73 and A96, but there is no distinctive general trend. The significant descriptives from figs. 6.9a and 6.9b show that

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	559	38.44	19.04	36.0	2	100	98
A6	127	46.99	24.06	44.0	9	100	91
A9	466	43.93	18.49	41.0	6	100	94
A70	31	50.29	25.00	44.0	9	92	83
A96	155	51.94	21.38	54.0	2	100	98
A7	130	53.46	24.19	54.0	6	100	94
A73	129	46.49	22.88	42.0	7	100	93
A99	116	33.21	18.03	31.0	5	85	80
A92	66	53.15	21.80	52.5	14	98	84
A93	21	40.81	17.72	37.0	13	70	57
A94	37	47.76	23.69	44.0	11	88	77

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.9: Group descriptives of *Street* and *Coverage*

the $\bar{x}$ of A3 is 10 % - 15 % lower than the $\bar{x}$ of A6, A7, A9, A73, A92 and A96. They also shows that the $\bar{x}$ of A6 is about 13 % higher than the $\bar{x}$ of A99. Therefore it can be interpreted that accidents on the A3, A99 are associated with significantly less dense jams than on A6, A7, A9, A73, A70, A92 and A96. The descriptives show also that the A7, A70, A92 and A96 are have considerable higher coverage, when the A3, A9, A93 and A99 have considerable lower coverage compared to the overall $\bar{x}$.

Kat

This section analyzes the correlated relations of the accident variable *Kat*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Kat* is shown in table 4.1. The Kruskal-Wallis test of *Kat* - *SAvg* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *SAvg* for the relations of *Kat* are present.

The Kruskal-Wallis test of the relation *Kat* - *TMax* produces a p-value of 0.0030, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TMax* between the groups of *Kat*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.8. It shows that all

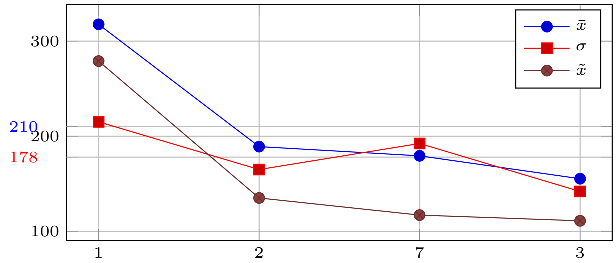
	1	2	3
2	0.00		
3	0.00	0.01	
7	0.00	0.02	0.78

Table 6.8: Pairwise Wilcoxon T-test for *Kat* and *Maximal Temporal Extent*

groups, besides of group 7 to group 3 have significant differences. The significant descriptives from

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	36	317.67	215.10	279	27	987	960
2	216	189.03	164.93	135	9	1257	1248
3	881	155.26	141.87	111	9	1323	1314
7	718	179.35	192.33	117	9	1341	1332

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.10: Group descriptives of *Kat* and *Maximal Temporal Extent*

figs. 6.10a and 6.10b present increasing $\bar{x}$ from lightly injured to deathly accident. Therefore it can be interpreted that the maximal temporal jam length significant increases with the gravity of the accident. Also the difference of group 1 to group 2, 3 and 7 is quite substantial, which means that the accidents with deaths are associated with 143 min longer jams, than others. The group of accidents with property damage does significantly differ from the others, but does not fit into the trend. A comparisons of $\bar{x}$ puts the property damage group in-between of accidents with slightly and heavily injured. The groups 2, 3 and 7 differ by 29 min on average. The Kruskal-Wallis test of the relation *Kat* - *TAvg* produces a p-value of 0.0004, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TAvg* between the groups of *Kat*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.9. It shows that all groups have significant

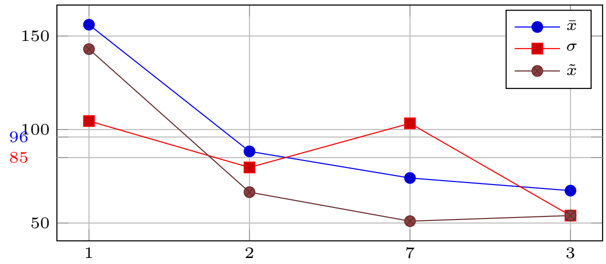
	1	2	3
2	0.00		
3	0.00	0.00	
7	0.00	0.00	0.05

Table 6.9: Pairwise Wilcoxon T-test for *Kat* and *Average Temporal Extent*

differences and therefore *TAvg* has a general trend. The descriptives from figs. 6.11a and 6.11b present increasing $\bar{x}$ from group 3 to 1. Therefore it can be interpreted that the average temporal jam length increases with the gravity of the accident. Also the difference of group 1 to group 2, 3 and 7 is quite substantial, which means that the accidents with deaths are associated with 80 min longer jams than others. The group of accidents with property damage does significantly

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	36	156.06	104.55	143.00	20	502	482
2	216	88.31	79.73	66.50	7	703	696
3	881	67.32	53.98	54.00	3	469	466
7	718	74.08	103.32	51.00	4	1326	1322

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.11: Group descriptives of *Kat* and *Average Temporal Extent*

differ from the others, but does not fit into the trend. A comparisons of $\tilde{x}$ puts the group in-between of accidents with slightly and heavily injured. The groups 2, 3 and 7 differ by 10 min on average. When comparing the $\bar{x}$ values of the maximal and average (temporal) extend (shown

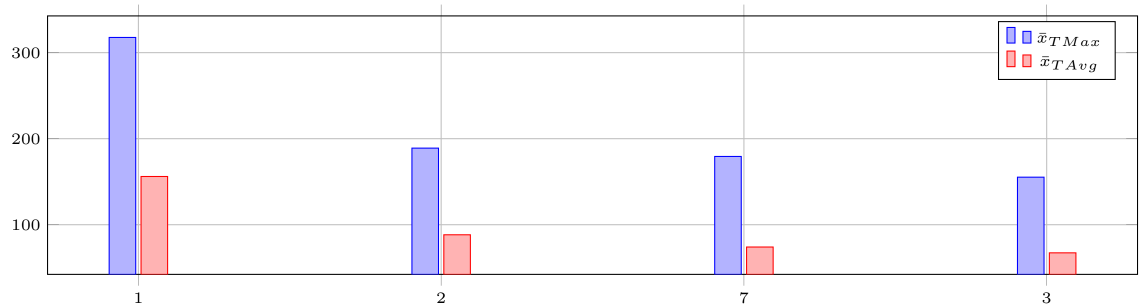


Figure 6.12: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ (*TMax/TAvg* by *Kat*)

in fig. 6.12) it becomes clear that the average is significantly lower than the maximum, which is to be expected. It also shows, that the difference between the groups are similar in the maximal and average extend. In can be described that the follow the same tend, that the jam duration significant increases with the gravity of the accident

The Kruskal-Wallis test of the relation *Kat* - *TDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TDist* between the groups of *Kat*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.10. It shows

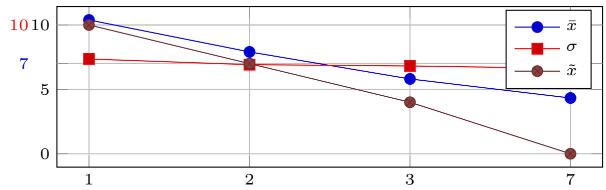
	1	2	3
2	0.05		
3	0.00	0.00	
7	0.00	0.00	0.00

Table 6.10: Pairwise Wilcoxon T-test for *Kat* and *Temporal Distance*

that all groups have significant differences. The significant descriptives from fig. 6.13a present increasing $\bar{x}$ from group 7 to 1. It can be interpreted that the temporal distance from accidents to jams increases by up to 6 min with the gravity of the accident.

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	36	10.39	7.36	10	0	22	22
2	216	7.91	6.92	7	0	24	24
3	881	5.81	6.82	4	0	24	24
7	718	4.33	6.63	0	0	24	24

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.13: Group descriptives of *Kat* and *Temporal Distance*

Typ

This section analyzes the correlated relations of the accident variable *Typ*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Typ* is shown in table 4.2. The Kruskal-Wallis test of *Typ* - *Cov* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Cov* for the relations of *Typ* are present.

The Kruskal-Wallis test of the relation *Typ* - *TDist* produces a p-value of 0.0003, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TDist* between the groups of *Typ*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.11. It show the

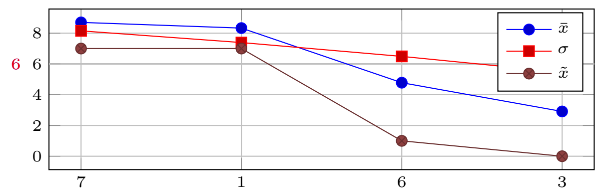
	1	3	4	5	6
3	0.00				
4	0.54	0.01			
5	0.54	0.58	0.23		
6	0.00	0.00	0.05	1.00	
7	1.00	0.00	0.54	0.54	0.00

Table 6.11: Pairwise Wilcoxon T-test for *Typ* and *Temporal Distance*

groups 3 and 6 differ significant from 1 and group 7 from 6. The group 4, 6 and 7 also differ significantly from 3, but there is no distinctive general trend. The significant descriptives in

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	300	8.33	7.39	7	0	24	24
3	120	2.91	5.46	0	0	22	22
6	1299	4.78	6.49	1	0	24	24
7	117	8.70	8.15	7	0	24	24

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.14: Group descriptives of *Typ* and *Temporal Distance*

figs. 6.14a and 6.14b show that the groups 1 and 7 have 5 min higher $\bar{x}$ on average than group 3 and 6. Therefore it can be interpreted that temporal distance of *driving* and *other* accidents (average of 8.5 min) is significantly higher than for *merging*, *crossing* and *straight traffic* accidents (average of 3.5 min).

UArt

This section analyzes the correlated relations of the accident variable *UArt*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *UArt* is shown in table 4.4. The Kruskal-Wallis test of *UArt1 - SAvg* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *SAvg* for the relations of *UArt1* are present.

The Kruskal-Wallis test of the relation *UArt1 - TDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TDist* between the groups of *UArt1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.12. It shows

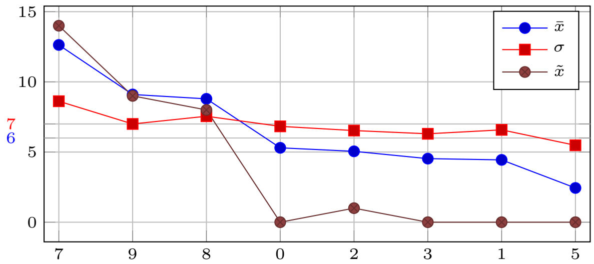
	0	1	2	3	4	5	6	7	8
5	0.04	0.17	0.00	0.00	0.01				
6	0.40	0.19	0.23	0.16	1.00	0.01			
7	0.00	0.00	0.00	0.00	1.00	0.00	1.00		
8	0.02	0.00	0.00	0.00	1.00	0.00	1.00	0.32	
9	0.01	0.00	0.00	0.00	1.00	0.00	1.00	0.50	1.00

Table 6.12: Pairwise Wilcoxon T-test for *UArt* and *Temporal Distance*, see table A.9 for complete table

that group 5 differs significantly from 2, 3 and 4. The groups 7, 8 and 9 differ significantly from the groups 0, 1, 2, 3 and 5. The descriptives in figs. 6.15a and 6.15b reveal that the

Group	n	$\bar{x}$	σ	$\tilde{x}$	*min*	*max*	Δ
0	63	5.30	6.83	0	0	22	22
1	86	4.44	6.58	0	0	24	24
2	831	5.05	6.53	1	0	24	24
3	447	4.53	6.30	0	0	24	24
5	90	2.44	5.48	0	0	24	24
7	35	12.63	8.62	14	0	24	24
8	165	8.79	7.55	8	0	24	24
9	124	9.10	7.00	9	0	24	24

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.15: Group descriptives of *UArt1* and *Temporal Distance*

groups 7, 8 and 9 have considerably higher $\bar{x}$, than the groups 1, 2, 3 and 5. Therefore it can be interpreted as accident collisions with *starting*, *standing*, *stopping*, *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles (average of 5 min) have a closer temporal reaction, when jams of accident collisions with *obstacles* or *left/right* nearby vehicles take longer to create (average of 8 min). Accidents of the category *turning* and *crossing* vehicles have the most immediate reaction of 2.5 min.

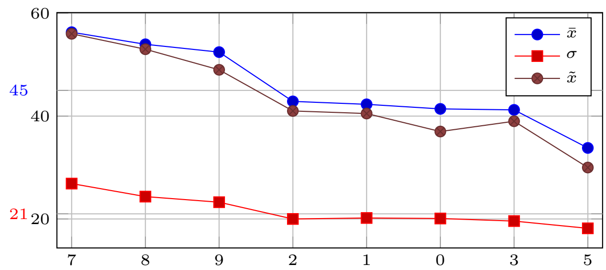
The Kruskal-Wallis test of the relation *UArt1 - Cov* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *Cov* between the groups of *UArt1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.13. It shows that group 5 differs significantly from 2, 3. Group 7 differs significantly from group 3 and group 8 from groups 0, 1 ,2 , 3 and 5. Also group 9 differs from 2, 3 and 4. The significantly

	0	1	2	3	4	5	6	7	8
5	0.41	0.10	0.00	0.01	0.65				
7	0.30	0.36	0.12	0.05	1.00	0.00	1.00		
8	0.01	0.01	0.00	0.00	1.00	0.00	1.00	1.00	
9	0.05	0.07	0.00	0.00	1.00	0.00	1.00	1.00	1.00

Table 6.13: Pairwise Wilcoxon T-test for *UArt1* and *Coverage*, see table A.10 for complete table

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	63	41.40	20.09	37.0	5	96	91
1	86	42.29	20.19	40.5	2	96	94
2	831	42.86	20.00	41.0	2	100	98
3	447	41.21	19.59	39.0	2	98	96
5	90	33.81	18.18	30.0	10	88	78
7	35	56.31	26.88	56.0	11	100	89
8	165	53.96	24.35	53.0	7	100	93
9	124	52.44	23.26	49.0	3	100	97

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.16: Group descriptives of *UArt* and *Coverage*

descriptives in figs. 6.16a and 6.16b show that the groups 7, 8 and 9 have considerable higher $\bar{x}$ than the groups 0, 1, 2, 3 and 5. Therefore it can be interpreted that accident collisions with *starting*, *standing*, *stopping*, *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles (average of 41 %) are associated with less dense jams than accident collisions with *obstacles* or *left/right* nearby vehicles (average of 54 %). Jams associated with accidents of the category *turning* and *crossing* vehicles are the least dense with 33 %.

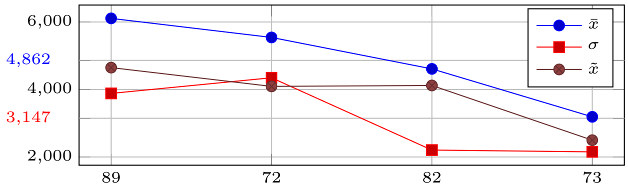
AUrs

This section analyzes the correlated relations of the accident variable *AUrs*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *AUrs* is shown in table 4.5. The variable value of group 0, which occurs in the dataset is undefined by BAYSIS and will be therefore neglected for the interpretation. The Kruskal-Wallis tests of *AUrs1-SDist* and *AUrs1-TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AUrs* for these relations are present.

The Kruskal-Wallis test of the relation *AUrs1 - SAvg* produces a p-value of 0.0228, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *SAvg* between the groups of *AUrs1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.11. It shows that only group 88 differs from 0 and 72. This is to be expected, because of the retained variable (see section 4.2). Since group 88 represents accident cause with *other obstacles* it can not be meaningful interpreted. The descriptives in figs. 6.17a and 6.17b can still be evaluated. They show that the (spatial) jam length increases with the accident causes *Slippery street due to rain* over *Visibility issues due to sun or glare* to *Slippery street due to snow or ice* by 2382 m on average.

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
72	41	5542.00	4347.10	4092	1009	16571	15562
73	95	3190.11	2152.19	2498	652	12288	11636
82	13	4610.15	2205.44	4119	1660	8426	6766
89	18	6106.22	3883.86	4648	347	12353	12006

(a) Table of all descriptives



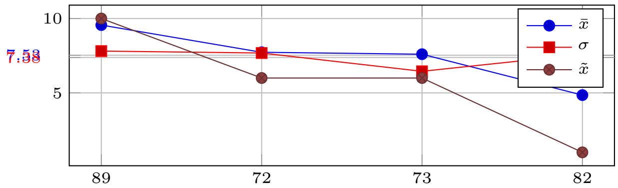
(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.17: Group descriptives of *AUrs* and *Average Spatial Extent*

The Kruskal-Wallis test of the relation *AUrs1 - TDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TDist* between the groups of *AUrs1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.12. It show that the only significant p-value is with group 0. Since the group is undefined from BAYSIS, there is no significant interpretation possible. The descriptives in figs. 6.18a and 6.18b can still

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
72	41	7.73	7.68	6	0	23	23
73	95	7.60	6.45	6	0	24	24
82	13	4.85	7.59	1	0	23	23
89	18	9.56	7.81	10	0	22	22

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

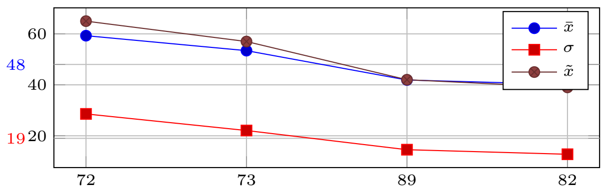
Figure 6.18: Group descriptives of *AUrs* and *Temporal Distance*

be evaluated. They show that the temporal distance between accident and congestion decreases with the accident causes *Slippery street due to snow or ice* and *Slippery street due to rain* (average of 7.7 min) to *Visibility issues due to sun or glare* to by 3 min on average.

The Kruskal-Wallis test of the relation *AUrs1 - Cov* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *Cov* between the groups of *AUrs1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.13. As with the variable *TDist*, it show that the only significant p-values are with group 0. Since the group is undefined from BAYSIS, there is no significant interpretation possible. The descriptives in

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
72	41	59.29	28.58	65.00	10	100	90
73	95	53.42	22.07	57.00	7	100	93
82	13	40.15	12.75	39.00	18	67	49
89	18	41.89	14.54	42.00	14	64	50

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.19: Group descriptives of *AUrs* and *Coverage*

figs. 6.19a and 6.19b can still be evaluated. They show that the jam coverage decreases with the accident causes *Slippery street due to snow or ice* over *Slippery street due to rain* to *Visibility* *issues due to sun or glare* to by 17 % on average.

AufHi

This section analyzes the correlated relations of the accident variable *AufHi*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *AufHi* is shown in table 4.6. The Kruskal-Wallis tests of *AufHi - TMax* and *AufHi - TAvg* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AufHi* for these relations are present.

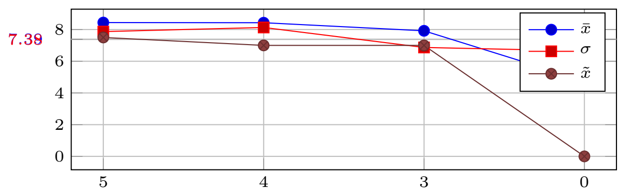
The Kruskal-Wallis test of the relation *AufHi - TDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TDist* between the groups of *AufHi*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.14. It shows, the groups of 3 and 4 differ significantly from 0. The significantly de-

	0	1	2	3	4	5	8
3	0.00	1.00	1.00				
4	0.03	1.00	1.00	1.00			

Table 6.14: Pairwise Wilcoxon T-test for *AufHi* and *Temporal Distance*, see table A.14 for complete table

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	1404	4.77	6.65	0.0	0	24	24
3	380	7.92	6.87	7.0	0	24	24
4	44	8.43	8.13	7.0	0	24	24
5	16	8.44	7.87	7.5	0	21	21

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.20: Group descriptives of *AufHi* and *Temporal Distance*

scriptives in figs. 6.20a and 6.20b show that accidents with *trees* ($\bar{x}$ of 4.7 min) have a much shorter distance to the adjacent jam, than accidents with the *guardrail* or *other obstacles* ($\bar{x}$ of 8.4 min).

The Kruskal-Wallis test of the relation *AufHi - Cov* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *Cov* between the groups of *AufHi*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.15. It shows

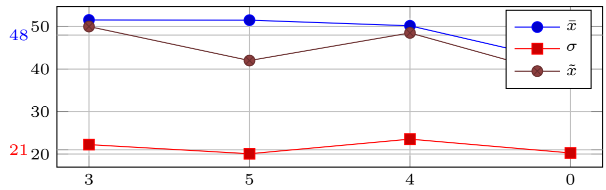
	0	1	2	3	4	5	8
3	0.00	1.00	1.00				

Table 6.15: Pairwise Wilcoxon T-test for *AufHi* and *Coverage*, see table A.15 for complete table

that only groups of 3 differs significantly from 0. The significantly descriptives in figs. 6.21a

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	1404	41.49	20.26	38.00	2	100	98
3	380	51.55	22.22	50.00	3	100	97
4	44	50.18	23.51	48.50	10	100	90
5	16	51.50	20.07	42.00	22	89	67

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.21: Group descriptives of *AufHi* and *Coverage*

and 6.21b show that accidents collisions with *trees* are associated with 10 % less dense jams, than accidents with the *guardrail* or *other obstacles*.

Lich

This section analyzes the correlated relations of the accident variable *Lich*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Lich* is shown in table 4.8.

The Kruskal-Wallis test of the relation *Lich1* - *Cov* produces a p-value of 0.0021, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *Lich1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.16. It shows that

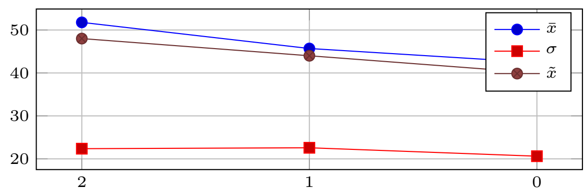
	0	1
1	0.19	
2	0.00	0.05

Table 6.16: Pairwise Wilcoxon T-test for *Lich1* and *Coverage*

only group 2 differs significantly from group 0. With the descriptives from figs. 6.22a and 6.22b

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	1492	42.42	20.62	40	2	100	95
1	97	45.69	22.58	44	10	100	90
2	262	51.77	22.37	48	3	100	93

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.22: Group descriptives of *Lich1* and *Coverage*

it can be stated that accidents during *darkness* are associated with 10 % denser jams than in *daylight*.

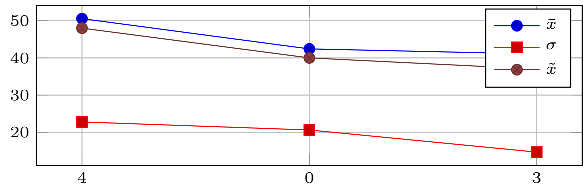
The Kruskal-Wallis test of the relation *Lich2* - *Cov* produces a p-value of 0.0018, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *Lich2*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.17. Similar to *Lich1* the table 6.17 only group 4 differs significantly from group 0. With the descriptives from

	0	3
3	0.98	
4	0.00	0.18

Table 6.17: Pairwise Wilcoxon T-test for *Lich2* and *Coverage*

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	1492	42.42	20.62	40	2	100	95
3	16	41.06	14.67	37	18	75	57
4	343	50.55	22.78	48	3	100	93

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.23: Group descriptives of *Lich2* and *Coverage*

figs. 6.23a and 6.23b it can state like for that accidents in ares where street light are disabled are associated with 8 % denser jams than in daylight, like *Lich1*.

Zust

This section analyzes the correlated relations of the accident variable *Zust*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Zust* is shown in table 4.9.

The Kruskal-Wallis test of the relation *Zust1* - *Cov* produces a p-value of 0.0013, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *Zust1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.18. It shows that

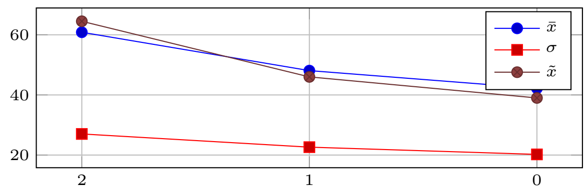
	0	1
1	0.00	
2	0.00	0.00

Table 6.18: Pairwise Wilcoxon T-test for *Zust* and *Coverage*

all groups differ significantly. This means there is a general trend, which can be interpreted with

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	1398	42.19	20.22	39.00	2	100	94
1	413	48.09	22.63	46.00	3	100	95
2	40	60.85	27.02	64.50	13	100	87

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.24: Group descriptives of *Zust1* and *Coverage*

figs. 6.24a and 6.24b. According to the descriptive the jam coverage increased with accidents in *dry* - *wet* - *ice* road conditions by 20 %.

WoTag

This section analyzes the correlated relations of the accident variable *baysis*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered.WoTag

The Kruskal-Wallis test of the relation *WoTag* - *Cov* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *Cov* between the groups of *WoTag*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.19. It show

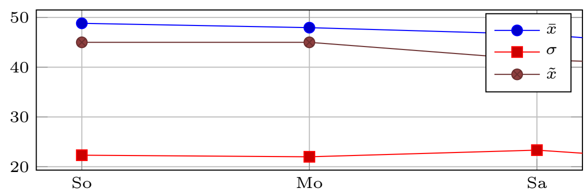
	Di	Mi	Do	Fr	Sa	So	Mo
So	0.00	0.02	0.00	0.14	1.00		
Mo	0.00	0.01	0.00	0.10	1.00	1.00	

Table 6.19: Pairwise Wilcoxon T-test for *WoTag* and *Coverage*, see table A.16 for complete table

that the groups *So* and *Mo* differ from groups *Di*, *Mi*, *Do* and *Fr*. With the descriptives in

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Mo	256	47.95	21.99	45.00	2	100	98
Di	280	40.62	20.88	37.00	6	100	94
Mi	287	41.47	18.91	39.00	6	90	84
Do	293	40.83	20.48	38.00	8	100	93
Fr	330	43.22	20.23	40.00	6	100	94
Sa	190	46.62	23.33	41.50	6	100	94
So	204	48.82	22.31	45.00	12	100	88

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.25: Group descriptives of *WoTag* and *Coverage*

fig. 6.25a we can interpret that accidents on Sundays and Mondays are associated with jams 8 % denser than on other week days.

Month

The Kruskal-Wallis test of *Zust* - *Cov* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Cov* for the relations of *Zust* are present.

6.2.2 Congestion - Accidents categorizes as Jam Initiator

The correlation matrix table for the congestion - accident dataset, which are classified as *Jam Initiator* (see table A.17) is visual presented in fig. 6.26 showing the the correlation of each variable combination. When visual analyzing fig. 6.2 and checking the guidelines for a strong correlation in reference to the applied coefficient (identifiable with table A.19) we get a list of strongly correlated variable combinations (see table 6.20). Since the focus of the thesis are the correlations between accidents and jams, these are only collected from the bottom-left rectangle of the matrix, where the congestion and accidents variables intersect. Correlations of the kind congestion - congestion or accident - accident are not considered. Next we need to verify that the

Category	*Strong*
Strasse	TMax, TAvg, SMax, SAvg, Cov, TLCar
Kat	TMax, TAvg, SMax, SAvg
Typ	SAvg, TDist, Cov
UArt1	TMax, TAvg, SMax, SAvg, TDist, Cov, TLCar
AUrs1	TMax, TAvg, SMax, SAvg, TDist, Cov, TLHGV
AUrs2	TMax, TAvg, SAvg, TDist
AufHi	TMax, TAvg, TDist, Cov
Char1	TDist
Lich1	TDist
Lich2	TDist
Zust1	Cov
Zust2	TAvg, SAvg
Month	SMax, Cov, TLHGV

Table 6.20: List of incident variables and their strong correlated congestion variable from the congestion - accident matched data which are classified as *Jam Initiator*

correlation is significant and what the correlation predicates. Therefore each correlation will be evaluated with the Post Hoc test, defined in section 3.6. In the following sections, the correlated relations of the variables in table 6.20 are analyzed and an initial interpretation of each significant correlation is introduced. Groups with an insufficient sample size (see section 3.5.1 are neglected and not shown to reduce the overall table size. The descriptive tables, showing the count (n), mean ($\bar{x}$), standard deviation (σ), median ($\tilde{x}$), min, max and range (Δ) therefore only contain groups with significant sample sizes.

Strasse

The Kruskal-Wallis tests of *Strasse - TMax*, *Strasse - TAvg*, *Strasse - SMax*, *Strasse - SAvg*, *Strasse - Cov* and *Strasse-TLCar* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Kat* for these relations are present.

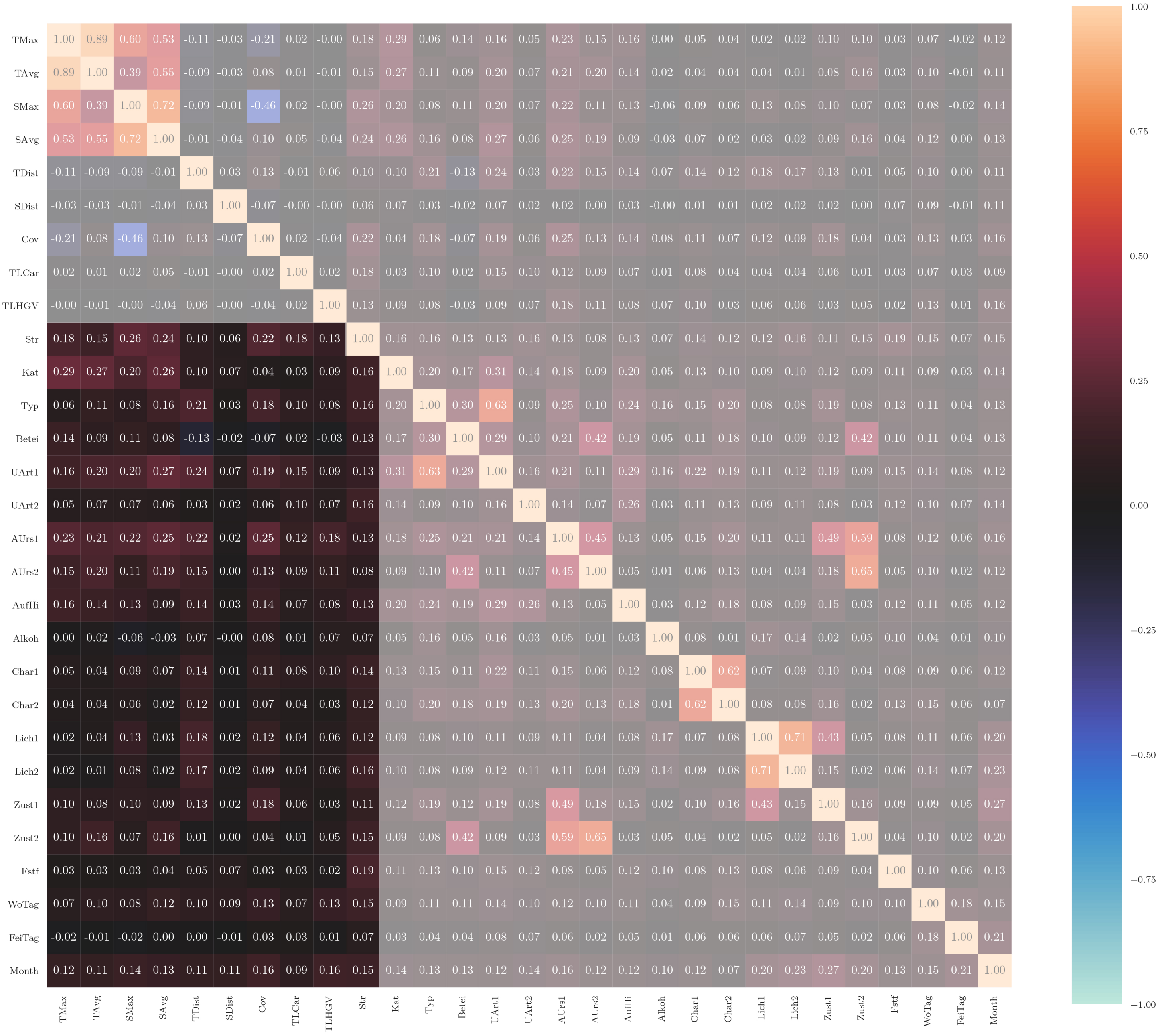


Figure 6.26: Correlation matrix for congestion-accident matched data classified as *Jam Initiator* and calculated with V, η, τ, r_{pq}, r

Kat

This section analyzes the correlated relations of the accident variable *Kat*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Kat* is shown in table 4.1. The Kruskal-Wallis tests of *Kat*-*SMax* and *Kat*-*SAvg* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Kat* for these relations are present.

The Kruskal-Wallis test of the relation *Kat* - *TMax* produces a p-value of 0.0001, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TMax* between the groups of *Kat*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.21. It shows that

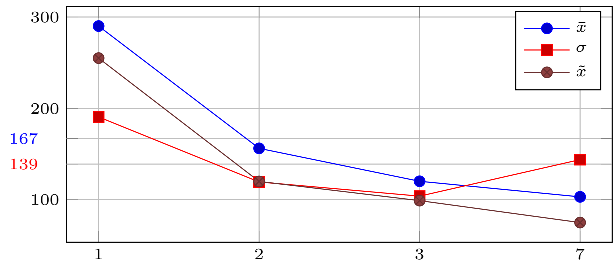
	1	2	3
2	0.00		
3	0.00	0.00	
7	0.00	0.00	0.00

Table 6.21: Pairwise Wilcoxon T-test for *Kat* and *Maximal Temporal Extent* (Jam Initiator)

all groups have significant differences and therefore *TMax* has a general trend. The significant

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	29	290.07	190.59	255.00	27	864	837
2	144	156.23	119.44	120.00	9	657	648
3	422	120.18	103.92	99.00	9	1116	1107
7	181	103.13	143.89	75.00	9	1341	1332

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.27: Group descriptives of *Kat* and *Maximal Temporal Extent* (Jam Initiator)

descriptives from figs. 6.27a and 6.27b present increasing means from accidents with property damage to deathly accidents. Therefore it can be interpreted that the maximal temporal jam length significant increases with the gravity of the accident. Also the difference of group 1 to group 2, 3 and 7 is considerable, which means that the accidents with deaths are associated with 164 min longer jams, than others. The groups 2, 3 and 7 differ on by 26 min on average. The Kruskal-Wallis test of the relation *Kat* - *TAvg* produces a p-value of 0.0087, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TAvg* between the groups of *Kat*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.22. It shows that all

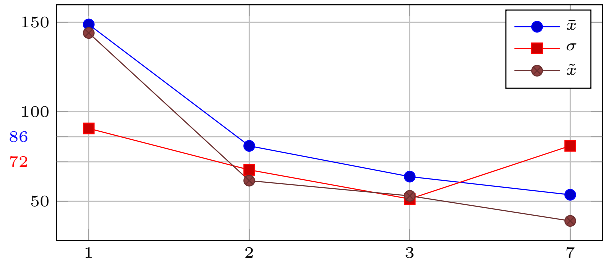
	1	2	3
2	0.00		
3	0.00	0.01	
7	0.00	0.00	0.00

Table 6.22: Pairwise Wilcoxon T-test for *Kat* and *Average Temporal Extent* (Jam Initiator)

groups have significant differences and therefore *TAvg* has a general trend. The significant descriptives from figs. 6.28a and 6.28b present increasing means from accidents with property damage to deathly accidents. Therefore it can be interpreted that the average temporal jam length significant increases with the gravity of the accident. Also the difference of group 1 to group 2, 3 and 7 is considerable, which means that the accidents with deaths are associated with 83 min longer jams, than others. The groups 2, 3 and 7 differ on by 10 min on average. When comparing the mean values of the maximal and average (temporal) extend (shown in fig. 6.29) it becomes clear that the average is significantly lower than the maximum, which is to be expected.

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	29	148.76	90.65	144.0	20	376	356
2	144	80.91	67.52	61.5	7	426	419
3	422	63.75	51.27	53.0	5	469	464
7	181	53.61	80.96	39.0	4	920	916

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.28: Group descriptives of *Kat* and *Average temporal Extent* (Jam Initiator)

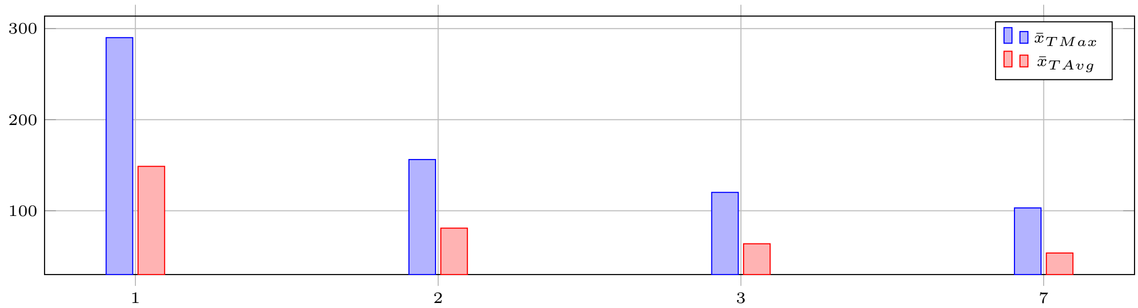


Figure 6.29: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ (*TMax/TAvg* by *Kat*) (Jam Initiator)

It also shows, that the difference between the groups are similar in the maximal and average extend. In can be described that they follow the same tend, that the jam duration significant increases with the gravity of the accident.

Typ

This section analyzes the correlated relations of the accident variable *Typ*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Typ* is shown in table 4.2. The Kruskal-Wallis test of *Typ* - *Cov* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Cov* for the relations of *Typ* are present.

The Kruskal-Wallis test of the relation *Typ* - *TDist* produces a p-value of 0.0264, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TDist* between the groups of *Typ*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.23. It shows that

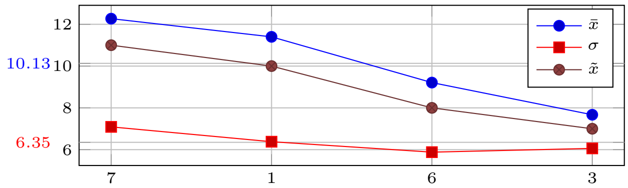
	1	3	4	5	6
3	0.05				
6	0.00	0.96	0.51	1.00	
7	1.00	0.04	1.00	0.96	0.01

Table 6.23: Pairwise Wilcoxon T-test for *Typ* and *Temporal Distance* (Jam Initiator), see table A.20 for complete table

group 6 differs significantly from group 1 and group 7 from 6, but there is no distinctive general trend. The significant descriptives in figs. 6.30a and 6.30b show that the groups 1 and 7 have

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
1	181	11.40	6.38	10.00	1	24	23
3	24	7.67	6.06	7.00	1	22	21
6	495	9.21	5.88	8.00	0	24	24
7	70	12.27	7.09	11.00	0	24	24

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.30: Group descriptives of *Kat* and *Temporal Distance* (Jam Initiator)

higher means than group 3 and 6. Therefore it can be interpreted that temporal distance of *driving* and *other* accidents (average of 12 min) is substantial higher than for *merging*, *crossing* and *straight traffic* accidents (average of 8 min).

UArt

This section analyzes the correlated relations of the accident variable *UArt*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *UArt* is shown in table 4.4. The Kruskal-Wallis tests of *UArt1* - *TAvg*, *UArt1* - *SMax*, *UArt1* - *SAvg* and *UArt1* - *TLCar* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *UArt* for these relations are present.

The Kruskal-Wallis test of the relation *UArt1* - *TMax* produces a p-value of 0.0014, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TMax* between the groups of *UArt1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.24. It shows that only

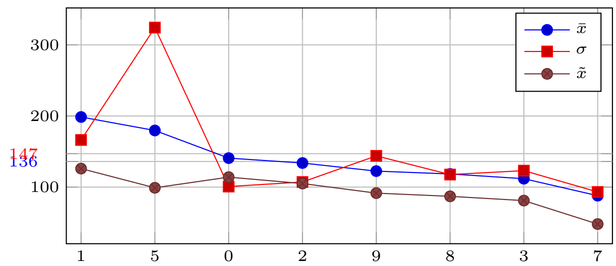
	0	1	2	3	4	5	6	7	8
3	1.00	0.04	0.00						
7	0.75	0.05	0.13	1.00	1.00	1.00	1.00		

Table 6.24: Pairwise Wilcoxon T-test for *UArt1* and *Maximal Temporal Extent* (Jam Initiator), see table A.21 for complete table

groups 3 and 7 differ significantly from group 1. The descriptives in figs. 6.32a and 6.32b reveal

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	24	140.75	100.75	114.0	30	375	345
1	31	198.58	166.25	126.0	39	612	573
2	344	133.84	107.15	105.0	9	864	855
3	144	111.79	123.13	81.0	9	1116	1107
5	15	179.60	324.31	99.0	15	1341	1326
7	23	87.91	93.15	48.0	12	384	372
8	107	118.46	117.62	87.0	18	813	79
9	80	122.47	143.91	91.5	12	1152	1140

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.31: Group descriptives of *UArt1* and *Maximal Temporal Extent* (Jam Initiator)

that $\bar{x}$ of group 3 and 7 differ by 24 min, but there is no considerable general trend. It can therefore only be interpreted that accidents with vehicles on the *separate lane in the same direction* can be associated with 24 min longer jams, than accidents with *obstacles*.

The Kruskal-Wallis test of the relation *UArt1 - TDist* produces a p-value of 0.0082, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TDist* between the groups of *UArt1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.25. The table shows,

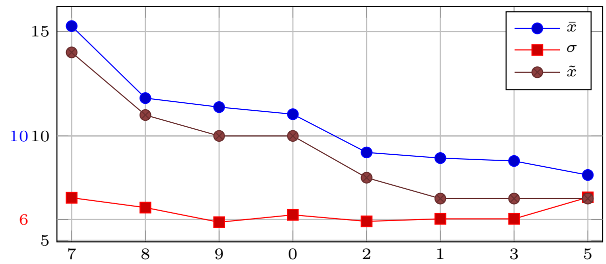
	0	1	2	3	4	5	6	7	8
7	1.00	0.06	0.00	0.00	1.00	0.10	1.00		
8	1.00	1.00	0.01	0.01	1.00	1.00	1.00	1.00	
9	1.00	1.00	0.05	0.01	1.00	0.78	1.00	0.87	1.00

Table 6.25: Pairwise Wilcoxon T-test for *UArt1* and *Temporal Distance* (Jam Initiator), see table A.22 for complete table

that groups 7, 8 and 9 differ from group 2 and 3. The descriptives in figs. 6.32a and 6.32b reveal

Group	n	$\bar{x}$	σ	$\tilde{x}$	*min*	*max*	Δ
0	24	11.04	6.22	10	2	22	20
1	31	8.94	6.03	7	0	24	24
2	344	9.21	5.91	8	1	24	23
3	144	8.80	6.03	7	1	24	23
5	15	8.13	7.06	7	0	22	22
7	23	15.26	7.04	14	4	24	20
8	107	11.81	6.57	11	1	24	23
9	80	11.38	5.87	10	2	24	22

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

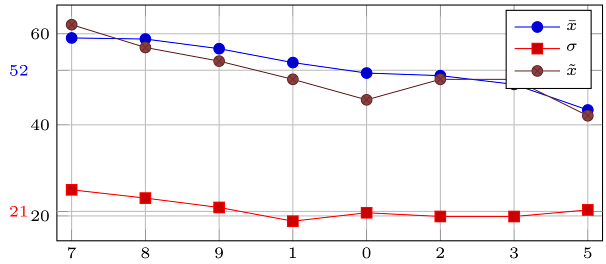
Figure 6.32: Group descriptives of *UArt1* and *Temporal Distance* (Jam Initiator)

that the groups 7, 8 and 9 have considerably higher $\bar{x}$, than the groups 1, 2, 3 and 5. Therefore it can be interpreted as accident collisions with *starting*, *standing*, *stopping*, *turning*, *crossing*, *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles (average of 9 min) have a closer temporal reaction, when jams of accident collisions with *obstacles* or *left/right* nearby vehicles take longer to create (average of 12 min).

The Kruskal-Wallis test of the relation *UArt1 - Cov* produces a p-value of 0.0337, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *UArt1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.23. It shows that the significant differences are not related to any specific group of *UArt1* and the correlation it not group specific. The descriptives in figs. 6.33a and 6.33b can still be evaluated. They show that the groups 7, 8 and 9 have considerable higher $\bar{x}$ than the groups 2, 3 and 5. Therefore it can be interpreted that accident collisions with *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles (average of 50 %) are associated with less dense jams than accident collisions with *obstacles* or *left/right* nearby vehicles (average of 58 %). Jams associated with accidents of the category *turning* and *crossing* vehicles are the least dense with 43 %.

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	24	51.38	20.70	45.5	23	96	73
1	31	53.68	18.85	50.0	28	96	68
2	344	50.81	19.88	50.0	7	100	93
3	144	48.90	19.87	50.0	5	98	93
5	15	43.27	21.32	42.0	12	88	76
7	23	59.09	25.74	62.0	18	100	82
8	107	58.85	23.92	57.0	7	100	93
9	80	56.73	21.86	54.0	18	100	82

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.33: Group descriptives of *UArt1* and *Coverage* (Jam Initiator)

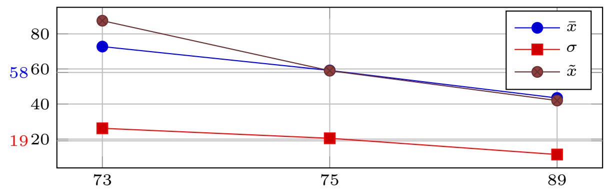
AUrs

This section analyzes the correlated relations of the accident variable *AUrs*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *AUrs* is shown in table 4.5. The variable value of group 0, which occurs in the dataset is undefined by BAYSIS and will be therefore neglected for the interpretation. The Kruskal-Wallis tests of *AUrs1-TMax*, *AUrs1-TAvg*, *AUrs1-SMax*, *AUrs1-SAvg*, *AUrs1-TDist* and *AUrs1-TLCar* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AUrs* for these relations are present.

The Kruskal-Wallis test of the relation *AUrs1 - Cov* produces a p-value of 0.0176, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *AUrs1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.24. It show shows that the significance differences is not group specific. The descriptives in figs. 6.34a and 6.34b can still be

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
73	20	72.75	26.16	87.5	20	100	80
75	63	59.16	20.47	59.0	15	100	85
89	11	43.45	11.22	42.0	25	62	37

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.34: Group descriptives of *AUrs* and *Temporal Distance* (Jam Initiator)

evaluated. They show that the jam coverage decreases with the accident causes *Slippery street due to rain* to *Cart track due to rain, snow or ice* to by 13 % on average.

All relations of the second *AUrs* variable *AUrs2* (*AUrs2-TMax*, *AUrs2-TAvg*, *AUrs2-SAvg* and *AUrs2-TDist*) produce p-values above the defined α-level. Therefore the null hypothesizes can't be rejected and no significant differences in *AUrs2* are present for these relations.

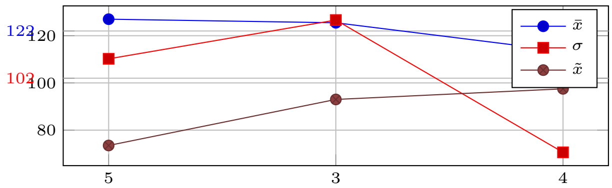
AufHi

This section analyzes the correlated relations of the accident variable *AufHi*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *AufHi* is shown in table 4.6. The Kruskal-Wallis tests of *AufHi-TAvg* and *AufHi-Cov* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AufHi* for these relations are present.

The Kruskal-Wallis test of the relation *AufHi - TMax* produces a p-value of 0.0411, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TMax* between the groups of *AufHi*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.25. It shows that the significant differences are not related to any specific group. The descriptives

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
3	241	125.49	126.64	93.0	12	1152	1140
4	24	113.50	70.60	97.5	21	321	300
5	12	127.00	110.21	73.5	45	405	360

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

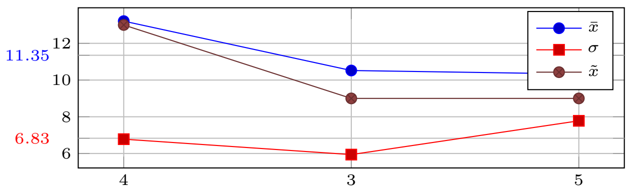
Figure 6.35: Group descriptives of *AufHi* and *Maximal Temporal Extent* (Jam Initiator)

in figs. 6.35a and 6.35b can still be evaluated. They show $\bar{x}$ 12 min lower for accident *Other objects*, than for *Guardrail* and *No collision*. This doesn't provide a meaningful base for interpretation.

The Kruskal-Wallis test of the relation *AufHi - TDist* produces a p-value of 0.0007, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TDist* between the groups of *AufHi*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.26. It shows that the significant differences are not related to any specific group. The descriptives

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
3	241	10.52	5.94	9	1	24	23
4	24	13.21	6.78	13	2	24	22
5	12	10.33	7.78	9	1	21	20

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.36: Group descriptives of *AufHi* and *Temporal Distance* (Jam Initiator)

in figs. 6.36a and 6.36b can still be evaluated. They show $\bar{x}$ 3 min higher for accident *Other objects*, than for *Guardrail* and *No collision*. This doesn't provide a meaningful base for interpretation.

Char

The Kruskal-Wallis test of *Char* - *TDist* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *TDist* for the relations of *Char* are present.

Lich

The Kruskal-Wallis tests of *Lich1* - *TDist* and *Lich2* - *TDist* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Lich* for these relations are present.

Zust

This section analyzes the correlated relations of the accident variable *Zust*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The encoding and description of the variable *Zust* is shown in table 4.9. The Kruskal-Wallis tests of *Zust2* - *SAvg* and *Zust2* - *TAvg* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Zust2* for these relations are present.

The Kruskal-Wallis test of the relation *Zust1* - *Cov* produces a p-value of 0.0046, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *Zust1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.26. The table

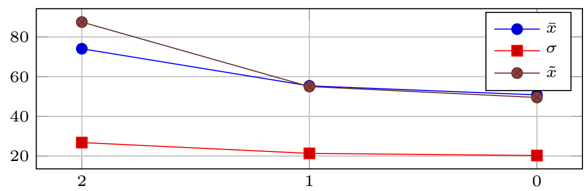
	0	1
0		
1	0.04	
2	0.00	0.01

Table 6.26: Pairwise Wilcoxon T-test for *Zust1* and *Coverage* (Jam Initiator)

shows that all groups differ significantly, which means that there is a general trend. With the

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
0	548	50.80	20.28	49.50	6	100	94
1	208	55.35	21.35	55.00	5	100	95
2	18	74.06	26.80	87.50	18	100	82

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.37: Group descriptives of *Zust1* and *Coverage* (Jam Initiator)

descriptives in figs. 6.37a and 6.37b it can be interpreted that the coverage of the jam increases from *dry* over *wet* to *ice* by nearly 25 %.

Month

The Kruskal-Wallis tests of *Month* - *TMax*, *Month* - *Cov* and *Month* - *TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Zust* for these relations are present.

6.2.3 Congestion - Accidents categorizes as Jam Effector

The correlation matrix table for the congestion - accident dataset, which are classified as *Jam Effector* (see table A.17) is visual presented in fig. 6.48 showing the the correlation of each variable combination. When visual analyzing fig. 6.2 and checking the guidelines for a strong correlation in reference to the applied coefficient (identifiable with table A.19) we get a list of strongly correlated variable combinations (see table 6.27). Since the focus of the thesis are the correlations between accidents and jams, these are only collected from the bottom-left rectangle of the matrix, where the congestion and accidents variables intersect. Correlations of the kind congestion - congestion or accident - accident are not considered. Next we need to verify that the

Category	Strong
Strasse	TMax, TAvg, SMax, SAvg, Cov, TLHGV
Kat	TMax, TAvg, SAvg
UArt1	SAvg
AufHi	TMax, TAvg
WoTag	TAvg, SMax, Cov, TLHGV
Month	TMax, TAvg, SMax, SAvg, Cov, TLHGV

Table 6.27: List of incident variables and their strong correlated congestion variable from the congestion - accident matched data which are classified as *Jam Effector*

correlation is significant and what the correlation predicates. Therefore each correlation will be evaluated with the Post Hoc test, defined in section 3.6. In the following sections, the correlated relations of the variables in table 6.27 are analyzed and an initial interpretation of each significant correlation is introduced. Groups with an insufficient sample size (see section 3.5.1 are neglected and not shown to reduce the overall table size. The descriptive tables, showing the count (n), mean ($\bar{x}$), standard deviation (σ), median ($\tilde{x}$), min, max and range (Δ) therefore only contain groups with significant sample sizes.

Street

This section analyzes the correlated relations of the accident variable *Str*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The Kruskal-Wallis tests of *Strasse* - *SAvg* and *Strasse* - *TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Str* for these relations are present.

The Kruskal-Wallis test of the relation *Street* - *TMax* produces a p-value of 0.0104, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TMax* between the groups of *Street*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.28. The table shows that the roads A73, A94, A95 and A96 differ from the A3. The A73 also differs from the A9, but there is no distinctive general trend. The significant descriptives from figs. 6.39a and 6.39b that the mean of A3 is 140 min - 188 min higher than the means of A73, A94, A95 and A96. They also

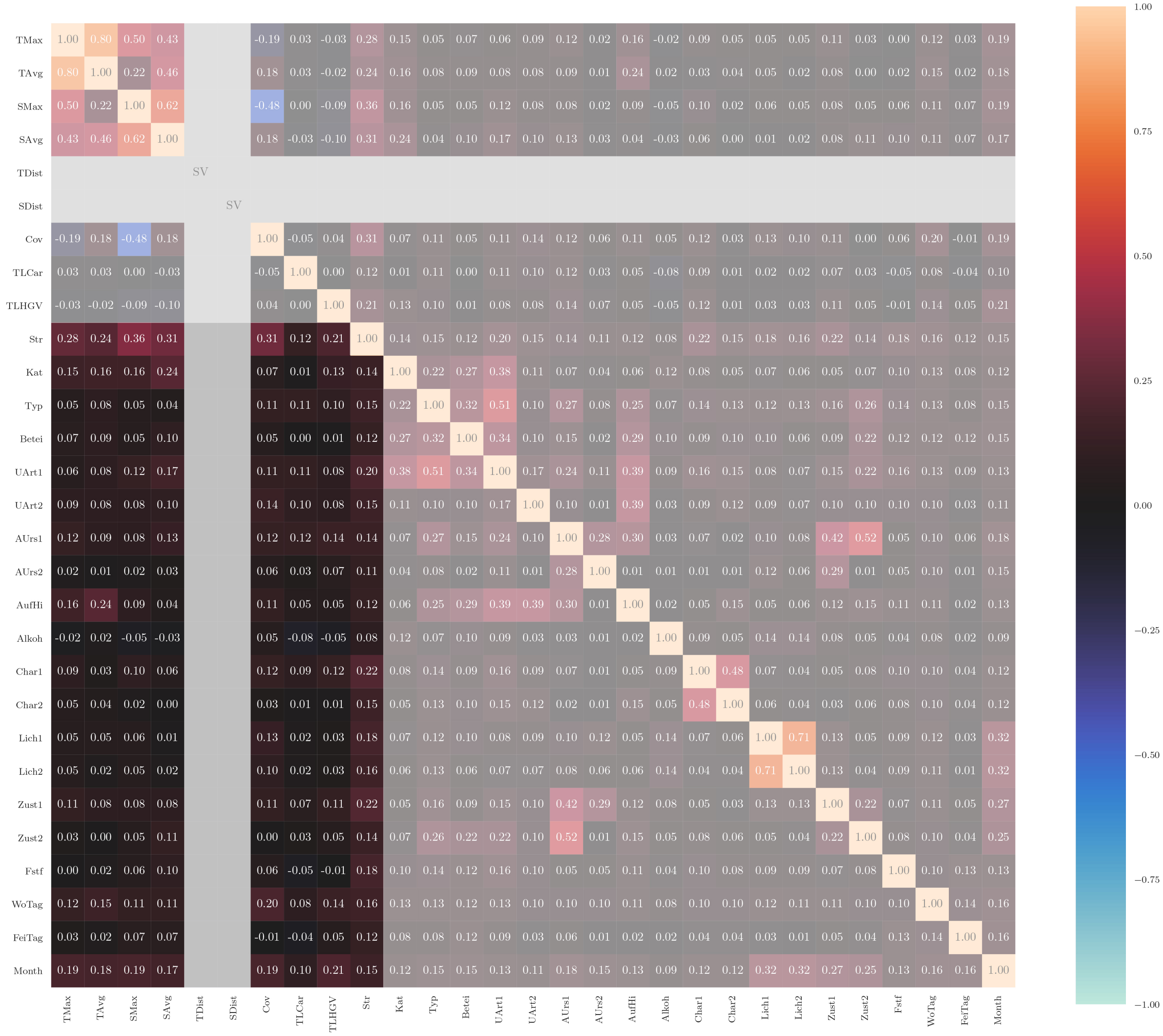


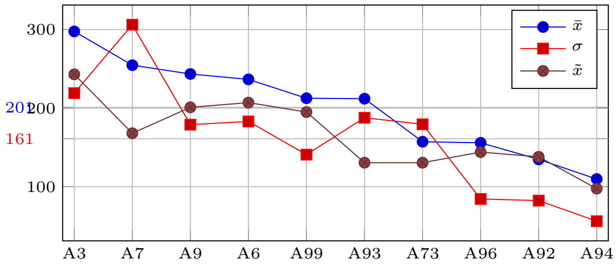
Figure 6.38: Correlation matrix for congestion-accident matched data classified as *Jam Effector* and calculated with V, η, τ, r_{pq}, r

	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A94	0.01	1.00	0.06	1.00	0.28	1.00							
A73	0.00	1.00	0.00	1.00	0.23	1.00	1.00	1.00					
A96	0.00	1.00	0.35	1.00	1.00	1.00	1.00	1.00	1.00				
A92	0.01	1.00	0.24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		

Table 6.28: Pairwise Wilcoxon T-test for *Street* and *Maximal Temporal Extent* (Jam Effector), see table A.30 for complete table

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	265	297.62	219.09	243.0	18	1257	1239
A6	37	236.59	183.03	207.0	18	705	687
A9	192	243.45	178.97	201.0	15	1194	1179
A99	63	212.57	140.78	195.0	21	681	660
A93	12	212.00	187.74	130.5	39	588	549
A94	14	109.71	56.07	97.5	42	249	207
A7	35	254.66	306.21	168.0	42	1341	1299
A73	52	157.21	179.42	130.5	18	1323	1305
A96	41	155.85	84.33	144.0	30	381	351
A92	21	134.71	82.25	138.0	18	354	336

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.39: Group descriptives of *Street* and *Maximal Temporal Extent* (Jam Effector)

show that the mean of A73 is 86 min lower than the mean of A9. Therefore it can be interpreted that accidents on the A3 and A9 are associated with significantly longer (temporal) jams than on the A73, A94, A95 and A96. The descriptives also show that the A73, A92, A94 and A96 are have considerable shorter durations, when the A3, A6, A7 and A9 have considerable longer durations compared to the general mean. The Kruskal-Wallis test of the relation *Street* - *TAvg* produces a p-value of 0.0003, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TAvg* between the groups of *Street*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.29. The table shows that the roads A99 and A73 differ significantly from

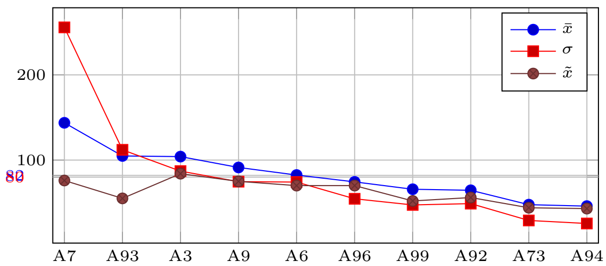
	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A99	0.02	1.00	1.00	1.00									
A73	0.00	1.00	0.00	1.00	1.00	1.00	1.00	0.02					

Table 6.29: Pairwise Wilcoxon T-test for *Strasse* and *Average Temporal Extent* (Jam Effector), see table A.31 for complete table

the A3. The A73 also differs significantly from A7. The significant descriptives from figs. 6.40a

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	265	104.05	87.02	84	7	703	696
A6	37	82.43	74.19	70	3	301	298
A9	192	91.31	74.64	75	5	575	570
A99	63	65.73	47.32	52	4	295	291
A93	12	104.83	112.03	55	7	343	336
A94	14	45.93	25.54	43	14	102	88
A7	35	143.74	255.78	76	15	1326	1311
A73	52	47.63	29.09	44	6	154	148
A96	41	74.39	54.53	70	6	247	241
A92	21	64.52	48.86	56	8	235	227

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.40: Group descriptives of *Street* and *Average Temporal Extent* (Jam Effector)

and 6.40b that the mean of A3 is 39 min - 57 min higher than the means of A99 and A73. They also show that the mean of A73 is 100 min lower than the mean of A7, which breaks the general trend of the variable and could be the result of errors. Never the less it can be interpreted that accidents on the A3 and A7 are associated with significantly longer (temporal) jams than on the A99 and A73. The descriptives also show that the A73, A92, A94 and A96 are have considerable shorter durations, when the A3, A7, A9 and A99 have considerable longer durations compared to the general mean. When comparing the mean values of the maximal and average (temporal)

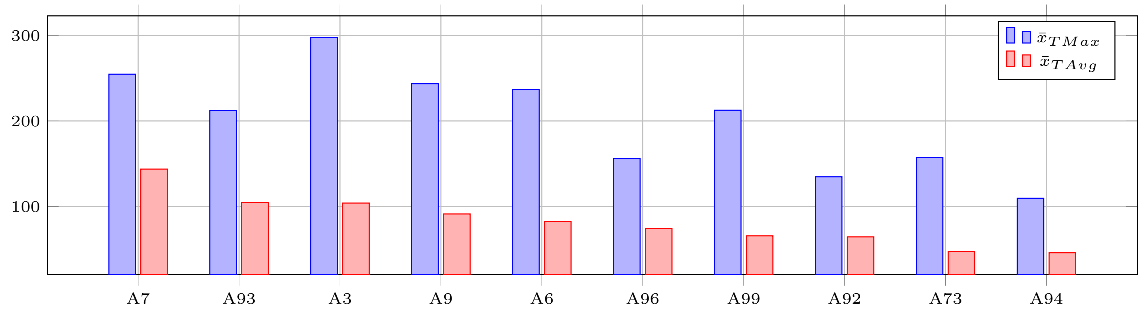


Figure 6.41: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ (*TMax*/*TAvg* by *Street*) (Jam Effector)

extend (shown in fig. 6.41) it becomes clear that the average variable has considerable lower values than the maximum variable, which is to be expected. It also shows, that the differences between the groups are mostly different in the maximal and average extend and vary considerably. In can be described that they follow tend, but A3, A9, A6 and A99 have higher maximal durations than the average trend.

The Kruskal-Wallis test of the relation *Street* - *SMax* produces a p-value of 0.0025, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *SMax* between the groups of *Street*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.30. The table shows

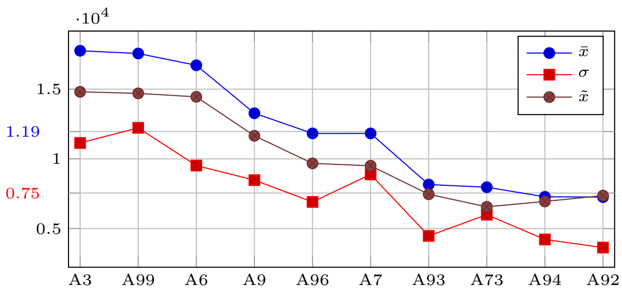
	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A9	0.00	1.00											
A93	0.07	0.13	1.00	1.00	1.00								
A94	0.01	0.03	0.41	1.00	0.24	1.00							
A7	0.08	0.65	1.00	1.00	1.00	1.00	1.00						
A73	0.00	0.00	0.00	1.00	0.00	1.00	1.00	0.95					
A96	0.13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.18				
A92	0.00	0.00	0.04	1.00	0.04	1.00	1.00	1.00	1.00	1.00	1.00		

Table 6.30: Pairwise Wilcoxon T-test for *Street* and *Maximal Spatial Extent* (Jam Effector), see table A.32 for complete table

that the roads of A73, A9, A92 and A94 differ significantly from A3. The roads A73, A92 and A93 also differ significantly from A6. The A73 and A92 also differ significantly from A9 and A99, but there is no distinctive trend. The significant descriptives from figs. 6.42a and 6.42b

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	265	17755.15	11139.22	14811.0	1566	46328	44762
A6	37	16711.43	9518.81	14449.0	2655	40033	37378
A9	192	13271.42	8473.71	11654.5	1315	49765	48450
A99	63	17558.70	12223.42	14698.0	2351	48278	45927
A93	12	8158.42	4473.23	7461.5	2896	16922	14026
A94	14	7277.64	4218.45	6947.0	1206	15550	14344
A7	35	11825.00	8885.99	9506.0	3041	43244	40203
A73	52	7964.98	5995.59	6559.5	1036	33764	32728
A96	41	11825.98	6911.19	9676.0	2006	27965	25959
A92	21	7255.76	3637.71	7364.0	1176	13522	12346

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.42: Group descriptives of *Street* and *Maximal Spatial Extent* (Jam Effector)

show that the mean of A3 is 4484 m - 10500 m higher than the means of A9, A73, A92 and A94.

They also shows that the groups A6 have a 9212 m higher mean on average than the groups A73 and A92. The mean of the groups A9 and A99 are 7805 m higher than the A73 and A92. Therefore it can be interpreted that accidents on the A3, A6, A9 and A99 are associated with significantly longer (spatial) jams than on A9, A73, A92 and A94. The descriptives show also that the A3, A6, A9 and A99 have a considerable longer lengths, when the A73, A92, A93 and A94 have considerable shorter lengths compared to the general mean.

The Kruskal-Wallis test of the relation *Street* - *Cov* produces a p-value of 0.0055, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *Street*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.31. The table shows

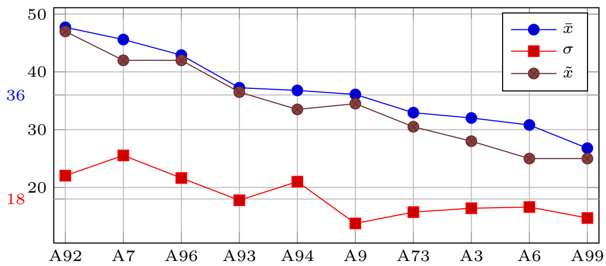
	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A9	0.01	1.00											
A99	1.00	1.00	0.00	1.00									
A7	0.25	1.00	1.00	1.00	0.02	1.00	1.00						
A96	0.21	1.00	1.00	1.00	0.02	1.00	1.00	1.00	1.00				
A92	0.03	0.43	1.00	1.00	0.01	1.00	1.00	1.00	0.61	1.00	1.00		

Table 6.31: Pairwise Wilcoxon T-test for *Street* and *Coverage* (Jam Effector), see table A.33 for complete table

that roads A9 and A92 differ significantly from A3. The road A99 differs significantly A9. The roads A7, A92 and A96 differ significantly from A99. The significant descriptives from figs. 6.43a

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A3	265	32.02	16.40	28.0	2	100	92
A6	37	30.81	16.61	25.0	9	65	56
A9	192	36.09	13.77	34.5	6	86	80
A99	63	26.79	14.71	25.0	7	63	56
A93	12	37.25	17.78	36.5	13	70	57
A94	14	36.79	21.00	33.5	11	77	66
A7	35	45.60	25.54	42.0	6	100	94
A73	52	32.94	15.73	30.5	7	77	70
A96	41	42.90	21.63	42.0	9	85	76
A92	21	47.71	22.04	47.0	21	88	67

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.43: Group descriptives of *Street* and *Coverage* (Jam Effector)

and 6.43b show that the mean of A3 is +4 % and -6 % different to the means of A9 and A92 respectively. They also shows that the mean of A9 is about 10 % higher than the mean of A99, which has a 16 % - 21 % lower mean on average than the A7, A92 and A96. Therefore it can be interpreted that accidents on the A3 and A99 are associated with significantly less dense jams than on A7, A9, A92 and A96. The descriptives show also that the A7, A92 and A96 are have considerable higher coverage, when the A3, A6, A73 and A99 have considerable lower coverage compared to the general mean.

Kat

The Kruskal-Wallis tests of *Kat* - *TMax*, *Kat* - *TAvg* and *Kat* - *SAvg* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Kat* for these relations are present.

UArt

The Kruskal-Wallis test of *UArt1 - SAvg* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *SAvg* for the relations of *UArt1* are present.

AufHi

Both Kruskal-Wallis tests of *AufHi - TMax* and *AufHi - TAvg* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AufHi* for Both relations are present.

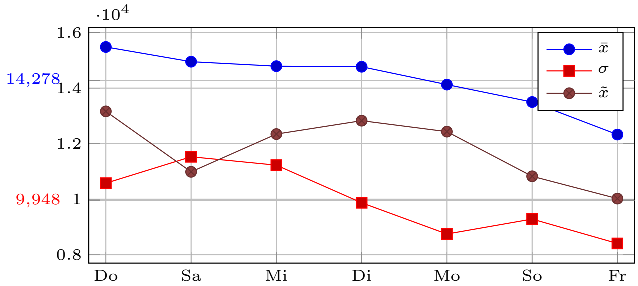
WoTag

This section analyzes the correlated relations of the accident variable *WoTag*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. Both Kruskal-Wallis tests of *WoTag - TAvg* and *WoTag - Cov* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *WoTag* for Both relations are present.

The Kruskal-Wallis test of the relation *WoTag - SMax* produces a p-value of 0.0175, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *SMax* between the groups of *WoTag*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.34. Although the Kruskal-Wallis test shows significant differences, the Wilcoxon table shows that the differences are not group specific. The descriptives in figs. 6.44a and 6.44b can still be interpreted. They

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Mo	115	14126.74	8746.49	12433	1315	38867	37552
Di	129	14769.57	9871.55	12827	2006	42736	40730
Mi	142	14790.29	11226.25	12345	1176	48278	47102
Do	124	15481.44	10576.67	13163	1772	42658	40886
Fr	89	12327.72	8406.29	10022	1036	49765	48729
Sa	65	14951.55	11527.80	10986	2402	46328	43926
So	73	13498.93	9284.69	10825	1524	42393	40869

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.44: Group descriptives of *WoTag* and *Maximal Spatial Extent* (Jam Effector)

show that accidents on Friday and Sunday can be associated with short jams, than on other weekdays.

The Kruskal-Wallis test of the relation *WoTag - TLHGV* produces a p-value of 0.0089, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *TLHGV* between the groups of *WoTag*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.32. The table show, that the groups of Wednesday and Sunday differ significantly from Tuesday and Wednesday

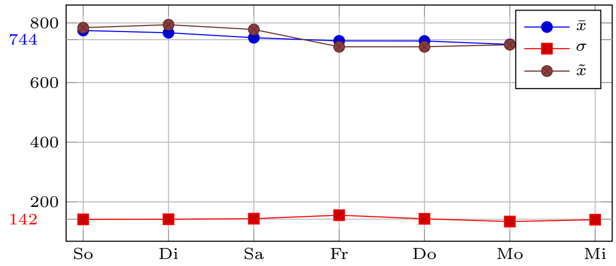
	Di	Do	Fr	Mi	Mo	Sa
Mi	0.03	1.00	1.00			
So	1.00	1.00	1.00	0.05	0.65	1.00

Table 6.32: Pairwise Wilcoxon T-test for *WoTag* and *Time-loss HGV* (Jam Effector), see table A.35 for complete table

respectively. With the descriptives in figs. 6.45a and 6.45b it can be interpreted, that the time

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Mo	89	728.10	134.08	727.00	514	986	472
Di	115	766.97	141.79	794.00	518	999	481
Mi	124	709.44	140.31	692.50	501	995	494
Do	129	739.48	143.28	720.00	511	983	472
Fr	142	739.78	155.44	720.00	502	998	496
Sa	65	750.46	143.80	778.00	507	988	481
So	73	774.67	141.34	784.00	534	994	460

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.45: Group descriptives of *WoTag* and *Time-loss Heavy goods vehicles* (Jam Effector)

loss for heavy goods vehicles on Tuesdays and Sundays (only theoretical, since there is no HGV traffic allow in Sundays) is 70 hours higher that on Wednesdays.

Month

This section analyzes the correlated relations of the accident variable *Month*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The Kruskal-Wallis tests of *Month* - *TMax*, *Month* - *TAvg*, *Month* - *TLHGV* and *Month* - *Cov* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Month* for these relations are present.

The Kruskal-Wallis test of the relation *Month* - *SMax* produces a p-value of 0.0208, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *SMax* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.33. The table

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Nov	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.15	0.01	1.00	

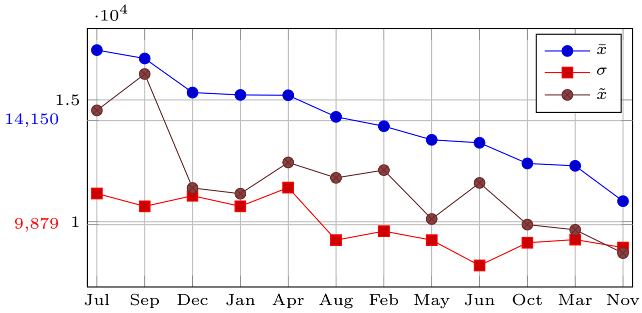
Table 6.33: Pairwise Wilcoxon T-test for *Month* and *Maximal Spatial Extent*, see table A.36 for complete table

shows, that only the groups of Nov differs significantly from Oct. The descriptives in fig. 6.46a show that the count n is distributed over the year with a peek in July and low in January. The November shows the shorts spatial length when the July show the longest length in $\bar{x}$ and σ. There is no distinctive general trend.

The Kruskal-Wallis test of the relation *Month* - *SAvg* produces a p-value of 0.0348, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	39	15204.44	10633.53	11151.0	1772	38867	37095
Feb	38	13921.39	9612.90	12118.5	1176	44548	43372
Mar	51	12295.18	9266.62	9669.0	1206	48278	47072
Apr	57	15190.47	11402.17	12433.0	1415	43070	41655
May	50	13361.66	9244.96	10112.5	1315	40805	39490
Jun	56	13241.27	8206.12	11594.0	2632	39685	37053
Jul	112	17045.40	11159.03	14569.5	1036	49765	48729
Aug	88	14304.86	9246.10	11801.5	2500	42393	39893
Sep	82	16700.20	10632.10	16063.5	2006	46328	44322
Oct	64	12392.00	9137.99	9882.5	1524	35354	33830
Nov	56	10844.89	8940.48	8712.0	1883	36151	34268
Dec	49	15302.61	11066.27	11386.0	1925	43244	41319

(a) Table of all descriptives



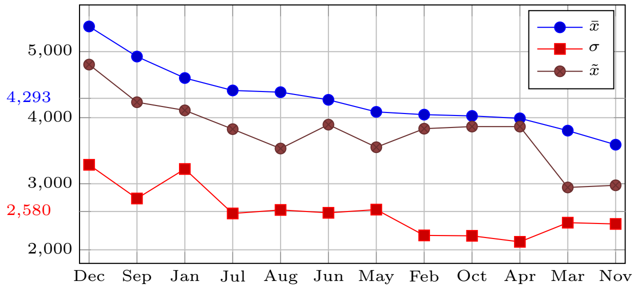
(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.46: Group descriptives of *Month* and *Maximal Spatial Extent* (Jam Effector)

significant differences in the variable *SAvg* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.37. Although the Kruskal-Wallis test shows significant differences, the Wilcoxon table shows that the differences are not group specific. The descriptives from fig. 6.47a show that November has the shortest

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	39	4600.31	3223.30	4111.0	829	14785	13956
Feb	38	4046.26	2217.96	3833.5	853	84260	7573
Mar	51	3803.96	2410.63	2944.0	747	10494	9747
Apr	57	3990.09	2120.45	3865.0	670	10320	9650
May	50	4087.96	2608.56	3552.5	393	10614	10221
Jun	56	4270.54	2559.92	3895.0	779	11206	10427
Jul	112	4412.27	2550.18	3826.5	784	15132	14348
Aug	88	4386.16	2603.28	3533.0	1036	13744	12708
Sep	82	4925.18	2776.80	4234.0	786	13605	12819
Oct	64	4026.47	2211.68	3865.5	358	8116	7758
Nov	56	3591.02	2391.39	2976.5	660	11167	10507
Dec	49	5380.88	3286.83	4804.0	1006	17805	16799

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.47: Group descriptives of *Month* and *Average Spatial Extent* (Jam Effector)

average spatial lengths when the December show the longest average lengths in $\bar{x}$ and σ. There is no distinctive general trend.

6.2.4 Congestion - Accidents categorizes as Jam Follower

The correlation matrix table for the congestion - accident dataset, which are classified as *Jam Follower* (see table A.17) is visual presented in fig. 6.48 showing the the correlation of each variable combination. When visual analyzing fig. 6.2 and checking the guidelines for a strong correlation in reference to the applied coefficient (identifiable with table A.19) we get a list of strongly correlated variable combinations (see table 6.34). Since the focus of the thesis are the correlations between accidents and jams, these are only collected from the bottom-left rectangle of the matrix, where the congestion and accidents variables intersect. Correlations of the kind congestion - congestion or accident - accident are not considered. Next we need to verify that the

Category	Strong
Strasse	TMax, TAvg, SMax, SAvg, TDist, SDist, Cov, TLHGV
Kat	TMax, SAvg
UArt1	TAvg, SAvg, TDist, Cov, TLHVG
UArt2	TDist
AUrs1	TDist, SDist, Cov, TLHGV
AufHi	TMax, TAvg, Cov
Lich1	Cov
WoTag	TAvg, SMax, SAvg, TDist, Cov, TLHGV
Month	TMax, TAvg, SMax, Cov, TLHGV

Table 6.34: List of incident variables and their strong correlated congestion variable from the congestion - accident matched data which are classified as *Jam Follower*

correlation is significant and what the correlation predicates. Therefore each correlation will be evaluated with the Post Hoc test, defined in section 3.6. In the following sections, the correlated relations of the variables in table 6.34 are analyzed and an initial interpretation of each significant correlation is introduced. Groups with an insufficient sample size (see section 3.5.1 are neglected and not shown to reduce the overall table size. The descriptive tables, showing the count (n), mean ($\bar{x}$), standard deviation (σ), median ($\tilde{x}$), min, max and range (Δ) therefore only contain groups with significant sample sizes.

Strasse

The Kruskal-Wallis tests of *Strasse - TMax*, *Strasse - TAvg*, *Strasse - SMax*, *Strasse - SAvg*, *Strasse - TDist*, *Strasse - SDist*, *Strasse - Cov* and *Strasse - TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Str* for these relations are present.

Kat

The Kruskal-Wallis tests of *Kat - TMax* and *Kat - SAvg* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Kat* for these relations are present.

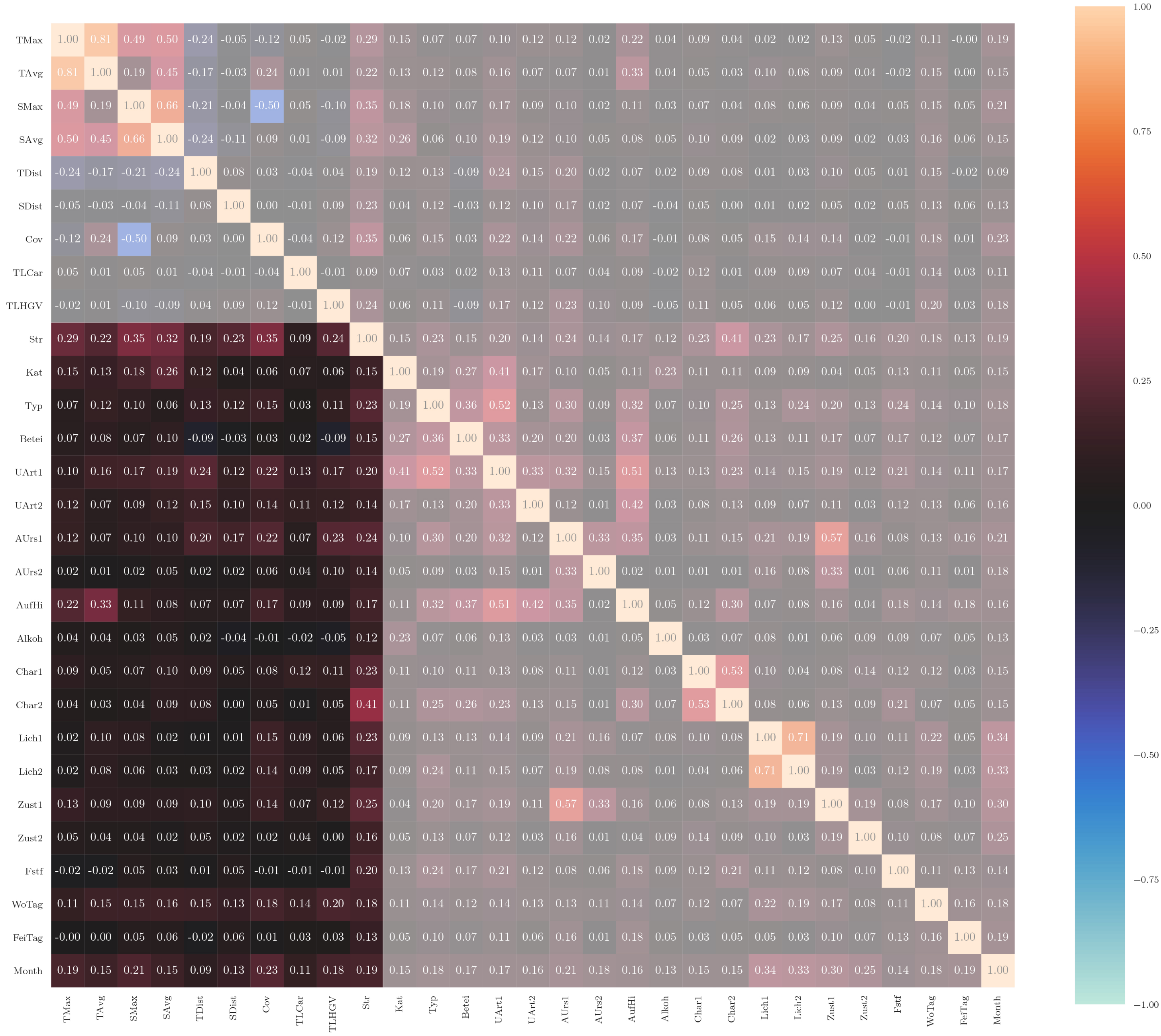


Figure 6.48: Correlation matrix for congestion-accident matched data classified as *Jam Effector* and calculated with V, η, τ, r_{pq}, r

UArt

The Kruskal-Wallis tests of *UArt1 - TAvg*, *UArt1 - SAvg*, *UArt1 - TDist*, *UArt1 - Cov*, *UArt1 - TLHGV* and *UArt2 - TDist* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *UArt* for these relations are present.

AUrs

This section analyzes the correlated relations of the accident variable *AUrs*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered. The Kruskal-Wallis tests of *AUrs1 - TDist* and *AUrs1 - TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AUrs* for these relations are present.

The Kruskal-Wallis test of the relation *AUrs1 - SDist* produces a p-value of 0.0372, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *SDist* between the groups of *AUrs1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.41. It show shows that the significance differences are group specific for group 73, but the descriptives reveal that all groups are uncertain, besides of group 73. Therefore the correlation can be interpreted and will be neglected.

The Kruskal-Wallis test of the relation *AUrs1 - Cov* produces a p-value of 0.0372, which is below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are significant differences in the variable *Cov* between the groups of *AUrs1*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table A.42. It show shows that the significance differences is not group specific and the descriptives further reveal that all groups are uncertain, besides of group 73. Therefore the correlation can be interpreted and will be neglected.

AufHi

The Kruskal-Wallis tests of *AufHi - TMax*, *AufHi - TAvg* and *AufHi - Cov* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *AufHi* for these relations are present.

Lich

The Kruskal-Wallis test of *Lich1 - Cov* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Cov* for the relations of *Lich1* are present.

WoTag

The Kruskal-Wallis tests of *WoTag - TMax*, *WoTag - SMax*, *WoTag - SAvg*, *WoTag - TDist*, *WoTag - Cov* and *WoTag - TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *WoTag* for these relations are present.

Month

The Kruskal-Wallis tests of *Month* - *TMax*, *Month* - *SMax*, *Month* - *SAvg*, *Month* - *Cov* and *Month* - *TLHGV* result in p-values above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *Month* for these relations are present.

6.2.5 Congestion - Roadworks in general

As with the congestion - accident datasets in sections 6.2.1 to 6.2.4 the congestion - roadwork dataset also needs to be analyzed for correlations. The correlation matrix table for the congestion-roadwork dataset (see table B.1) is visual presented in fig. 6.49 showing the the correlation of each variable combination. When visual analyzing fig. 6.49 and checking the guidelines for a strong correlation in reference to the applied coefficient (identifiable with table A.19) we get a list of strongly correlated variable combinations (see table 6.35). Since the focus of the thesis are the correlations between accidents and jams, these are only collected from the bottom-left rectangle of the matrix, where the congestion and accidents variables intersect. Correlations of the kind congestion - congestion or roadwork - roadwork are not considered. As table 6.35

Category	Strong
Strasse	TMax, TAvg, SMax, SAvg, TDist, SDist, Cov, TLCar, TLHGV
Month	TAvg, SMax, SAvg, TDist, SDist, Cov, TLCar

Table 6.35: List of incident variables and their strong correlated congestion variable from the congestion - roadwork matched data

show, only the variables Strasse and Month are correlated with congestion characteristics. Both variables aren't actually roadworks specific characteristics but more general features. They still will be evaluated to identify possible dependencies. For that it needs to be verified that the correlation is significant and what the correlation predicates. Therefore each correlation will be evaluated with the Post Hoc test, defined in section 3.6. In the following sections, the correlated relations of the variables in table 6.35 are analyzed and an initial interpretation of each significant correlation is introduced. Groups with an insufficient sample size (see section 3.5.1 are neglected and not shown to reduce the overall table size. The descriptive tables, showing the count (n), mean ($\bar{x}$), standard deviation (σ), median ($\tilde{x}$), min, max and range (Δ) therefore only contain groups with significant sample sizes.

Strasse

This section analyzes the correlated relations of the accident variable *Strasse*. Groups with an insufficient sample size (see section 3.5.1 are neglected and not considered.

The Kruskal-Wallis test of the relation *Strasse* - *TMax* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TMax* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.36. The table shows, that the groups A3 and A93 differ significantly from group A9, A7 and A6. The A93 also differs significantly from A96, A995 and A92. The A96 differs significantly from A3 and A99. The significant descriptives from figs. 6.50a and 6.50b show that the A3 has considerable longer maximal durations, when the A92, A96 and A995 have considerable shorter maximal durations compared to the overall $\bar{x}$. The maximal positive deviation from the overall $\bar{x}$ is the

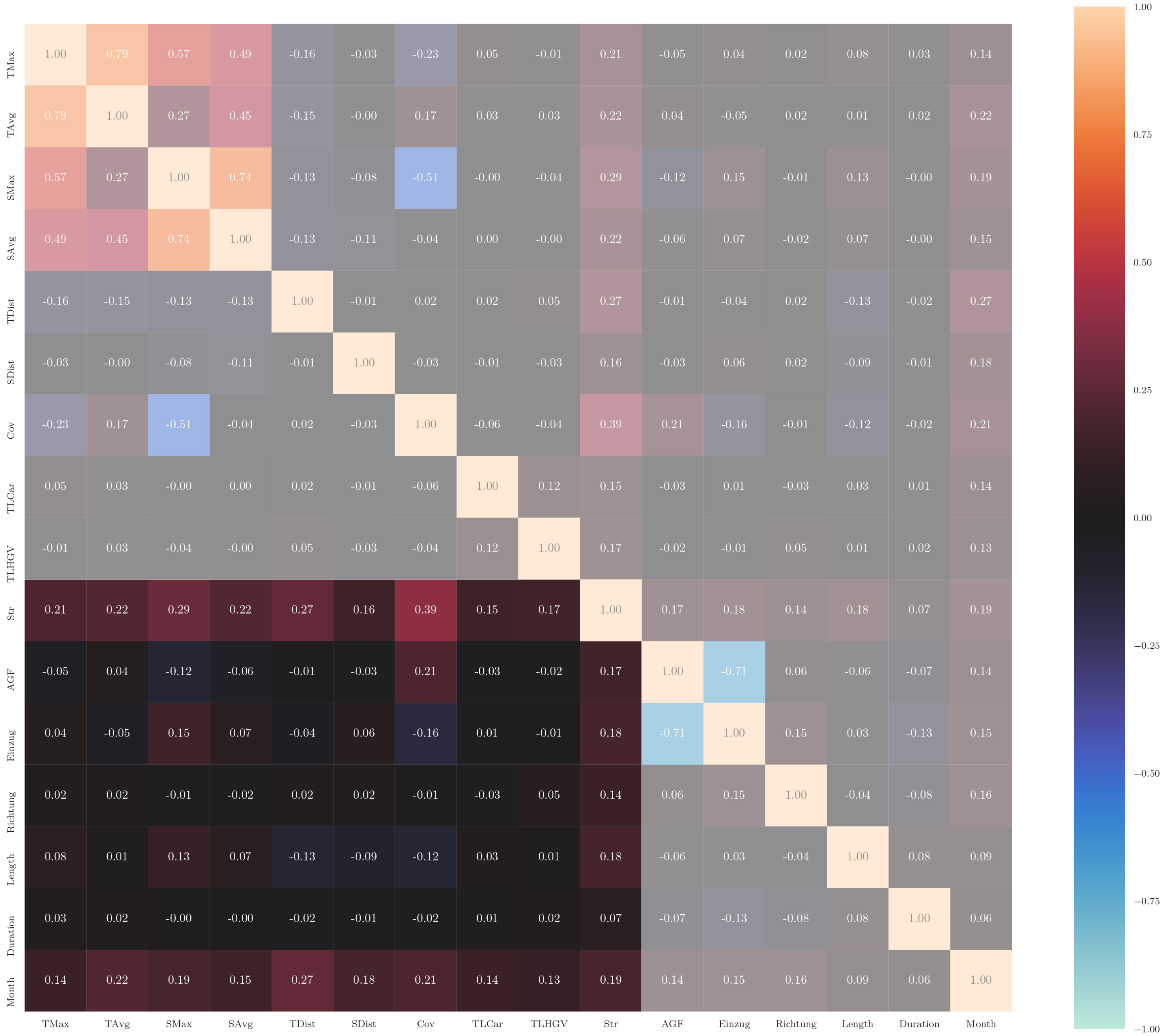


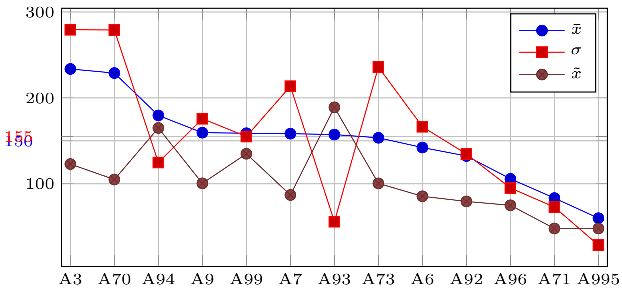
Figure 6.49: Correlation matrix for congestion-roadwork matched data, calculated with V, η, τ, r_{pq}, r

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A3	0.00	0.04	1.00	1.00	0.01	1.00										
A96	1.00	0.82	1.00	1.00	1.00	1.00	0.00	0.00								
A93	0.00	0.00	1.00	0.09	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00				
A94	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	0.01	0.48	1.00	1.00	1.00	1.00		

Table 6.36: Pairwise Wilcoxon T-test for *Strasse* and *Maximal Temporal Extent* (complete in table B.4)

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	159.55	175.78	100.50	9	1323	1314
A7	302	158.38	213.55	87.00	9	1323	1314
A70	50	228.90	279.03	105.00	15	963	948
A71	10	83.40	72.87	48.00	36	249	213
A6	198	142.24	166.46	85.50	9	864	855
A73	86	153.59	235.85	100.50	9	1323	1314
A3	1023	233.66	279.35	123.00	9	1326	1317
A99	312	158.88	155.05	135.00	9	1320	1311
A96	230	105.80	95.17	75.00	9	384	375
A995	14	60.00	28.70	48.00	18	105	87
A92	82	132.40	134.72	79.50	9	768	759
A93	160	157.33	55.97	189.00	9	312	303
A94	56	179.57	124.79	165.00	15	369	354

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.50: Group descriptives of *Strasse* and *Maximal Temporal Extent*

A3 with an increase of 88 min. The maximal negative deviation from the overall $\bar{x}$ is the A995 with an decrease of 90 min. The σ mostly follows the trend of $\bar{x}$, besides of A7, A73, A93 and A94. This means that $\bar{x}$ should be generally representable. In the cases of A7, A73, A93 and A94 additional analysis is needed for reliable results. The Kruskal-Wallis test of the relation *Strasse* - *TAvg* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TAvg* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.37. The table shows, that the groups A99

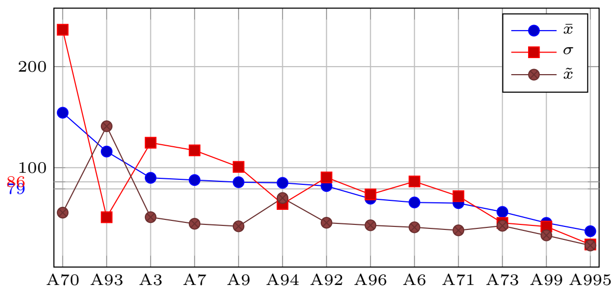
	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A99	0.00	0.00	0.08	1.00	1.00	1.00	0.00									
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.04	1.00	1.00						
A93	0.00	0.00	1.00	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
A94	1.00	1.00	1.00	1.00	0.14	1.00	1.00	0.00	0.63	0.53	1.00	1.00	0.02	1.00		

Table 6.37: Pairwise Wilcoxon T-test for *Strasse* and *Average Temporal Extent* (complete in table B.5)

and A93 differ significantly from group A9 and A7. The A93 also differs significantly from the A6, A73, A3, A99, A96, A995 and A92. The A92 differs significantly from A99 and A99 from A3. The A94 differs significantly from the A99 and A93. The significant descriptives from figs. 6.50a

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	85.61	100.78	42.00	5	575	570
A7	302	87.76	117.13	44.50	5	858	853
A70	50	154.38	236.49	55.50	6	966	960
A71	10	64.90	71.81	38.00	18	252	234
A6	198	65.60	86.36	41.00	4	630	626
A73	86	56.23	45.35	42.50	6	190	184
A3	1023	89.94	124.69	51.00	3	1326	1323
A99	312	45.38	41.73	33.00	3	291	288
A96	230	69.31	73.52	43.00	5	290	285
A995	14	37.14	23.98	23.00	8	74	66
A92	82	81.83	90.35	45.50	5	489	484
A93	160	115.96	50.96	141.00	6	199	193
A94	56	84.98	64.07	70.00	6	248	242

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.51: Group descriptives of *Strasse* and *Average Temporal Extent*

and 6.50b show that the A93 has considerable longer average durations, when the A92, A99 and A995 have considerable shorter average durations compared to the overall $\bar{x}$. The maximal positive deviation from the overall $\bar{x}$ is the A70 with an increase of 75 min. The maximal negative deviation from the overall $\bar{x}$ is the A995 with an decrease of 117 min. The σ follow the trend of $\bar{x}$ with a deviation at A70 and A93. This means that $\bar{x}$ should be generally representable. In the cases of A70 and A93 additional analysis is needed for reliable results. When comparing the $\bar{x}$

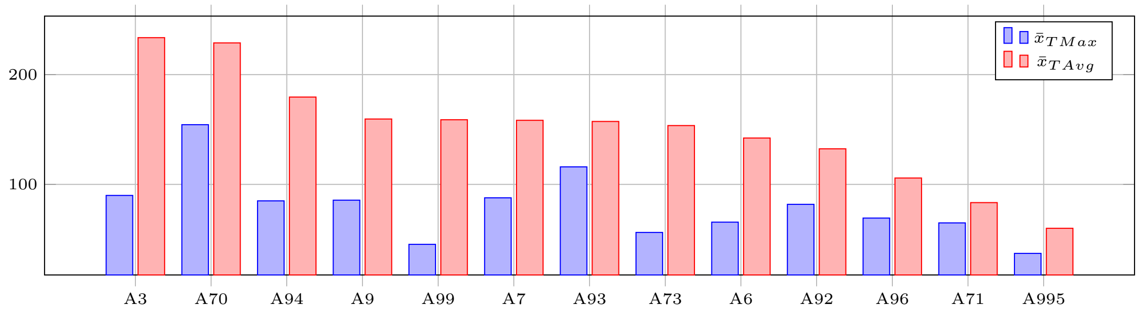


Figure 6.52: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ (*TMax/TAvg* by *Street*)

values of the maximal and average (temporal) extend (shown in fig. 6.52) it becomes clear that the average variable has considerable lower values than the maximum variable, which is to be expected. The distribution show a that there are considerable difference in the jam durations between the roads. There is a distinctive trend of magnitude in the maximal variable, but not in the same order as in the average. The Kruskal-Wallis test of the relation *Strasse* - *SMax* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SMax* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.38. The table shows, that the groups A3 and A93 differ significantly from

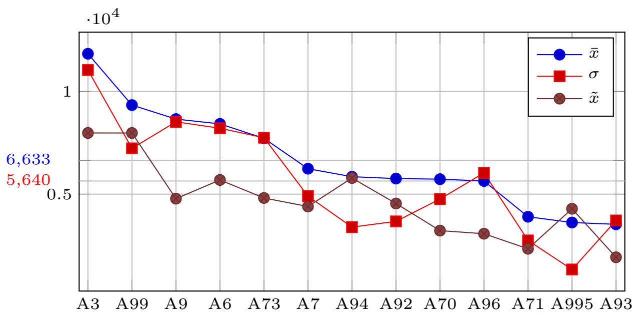
	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A3	0.00	0.00	0.00	0.52	0.01	0.05										
A96	0.00	0.06	1.00	1.00	0.00	1.00	0.00	0.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	0.11	0.03	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.04	1.00						
A93	0.00	0.00	0.02	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.73				
A94	1.00	1.00	1.00	1.00	1.00	1.00	0.04	0.04	0.89	1.00	1.00	1.00	0.00	1.00		

Table 6.38: Pairwise Wilcoxon T-test for *Strasse* and *Maximal Spatial Extent* (complete in table B.6)

group A6, A7, A70, A73, A9, A96, A92, A93 and A94. The group A99 also differs significantly from A96, A995, A92, A93 and A94. The A94 also differs significantly from A99. The significant

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	8655.57	8521.51	4778	1035	49765	48730
A7	302	6238.63	4902.30	4397	902	20030	19128
A70	50	5729.56	4763.20	3224	1365	20249	18884
A71	10	3899.00	2746.84	2339	2075	10225	8150
A6	198	8421.67	8209.85	5690	965	40033	39068
A73	86	7717.21	7747.18	4813	1095	33764	32669
A3	1023	11836.50	11045.29	7979	1014	47607	46593
A99	312	9337.49	7230.99	7978	991	48987	47996
A96	230	5639.41	6034.92	3072	951	27965	27014
A995	14	3618.14	1331.10	4288	1613	4825	3212
A92	82	5758.10	3673.73	4544	999	16931	15932
A93	160	3530.39	3718.54	1926	99	22528	21829
A94	56	5849.05	3394.57	5785	1025	12582	11557

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.53: Group descriptives of *Strasse* and *Maximal Spatial Extent*

descriptives from figs. 6.54a and 6.54b show that the A3, A6, A9 and A99 have considerable longer maximal lengths, when the A93 and A995 have considerable shorter maximal lengths compared to the overall $\bar{x}$. The maximal positive deviation from the overall $\bar{x}$ is the A3 with an increase of 5203 m. The maximal negative deviation from the overall $\bar{x}$ is the A93 with an decrease of 3103 m. The σ follow the trend of $\bar{x}$ with some deviation. This means that $\bar{x}$ should be generally representable. The Kruskal-Wallis test of the relation *Strasse* - *SAvg* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SAvg* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.39. It shows, that the groups A99, A96 and A93 differ significantly from

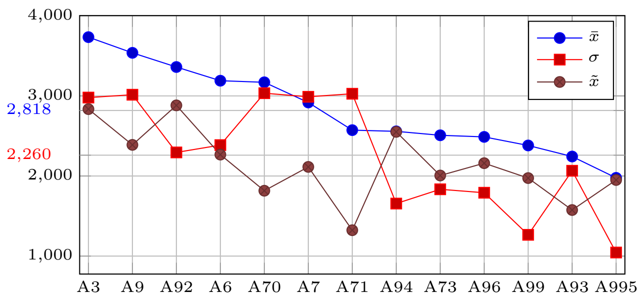
	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A3	1.00	0.00	1.00	1.00	1.00	0.02										
A99	0.01	1.00	1.00	1.00	0.75	1.00	0.00									
A96	0.00	1.00	1.00	1.00	0.44	1.00	0.00	1.00								
A93	0.00	0.13	1.00	1.00	0.00	1.00	0.00	0.11	1.00	1.00	0.00	1.00				

Table 6.39: Pairwise Wilcoxon T-test for *Strasse* and *Average Spatial Extent* (complete in table B.7)

group A9 and A3. The A93 also differs significantly from A92 and A73 from A3. The significant

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	3537.87	3014.80	2388.50	697	14785	14088
A7	302	2917.97	2988.99	2113.50	284	15602	15318
A70	50	3170.10	3033.98	1816.00	532	12543	12011
A71	10	2572.50	3027.05	1323.00	802	10425	9623
A6	198	3189.79	2385.78	2267.00	458	14150	13692
A73	86	2508.40	1834.03	2006.00	419	10039	9620
A3	1023	3733.90	2979.43	2835.00	355	15054	14699
A99	312	2381.29	1265.40	1974.00	502	5931	5429
A96	230	2488.22	1790.94	2160.00	404	9767	9363
A995	14	1976.93	1044.51	1950.00	569	3196	2627
A92	82	3360.88	2294.30	2882.00	457	11703	11246
A93	160	2243.26	2065.34	1575.00	461	11161	10700
A94	56	2557.62	1655.86	2551.00	506	6393	5887

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.54: Group descriptives of *Strasse* and *Average Spatial Extent*

descriptives from figs. 6.54a and 6.54b show that the A3, A9 and A92 have considerable longer average lengths, when the A93, A96 and A99 have considerable shorter average lengths compared to the overall $\bar{x}$. The maximal positive deviation from the overall $\bar{x}$ is the A3 with an increase of 915 m. The maximal negative deviation from the overall $\bar{x}$ is the A995 with an decrease of 842 m. The σ follows the trend of $\bar{x}$ with some heavy deviations. This means that $\bar{x}$ should be generally representable, but additional analysis is necessary for reliable result. When comparing the $\bar{x}$ values of the maximal and average (spatial) extend (shown in fig. 6.55) it becomes clear that the average variable has considerable lower values than the maximum variable, which is to be expected. The distribution shows a that there are considerable difference in the jam maximal lengths, but not as much in average lengths. There is a distinctive trend of magnitude in the maximal variable, but not in the same order as in the average.

The Kruskal-Wallis test of the relation *Strasse* - *TDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that

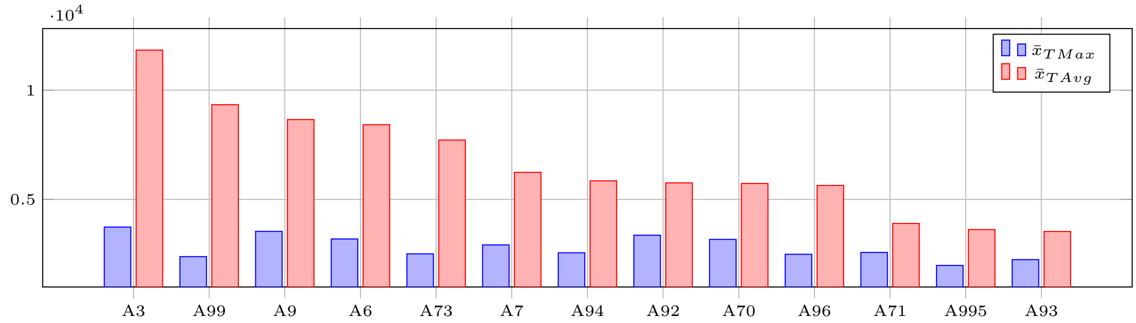


Figure 6.55: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ (*TMax/TAvg* by *Street*)

there are strong significant differences in the variable *TDist* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.40. It shows, that the groups A7, A3 and A92 differ significantly from group A9. The A96 differs

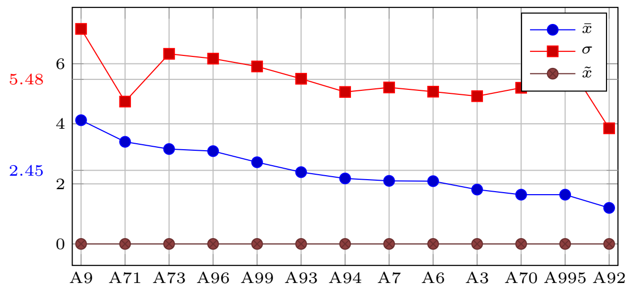
	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	0.00															
A3	0.00	1.00	1.00	1.00	1.00	0.85										
A96	1.00	1.00	1.00	1.00	1.00	1.00	0.03	1.00								
A92	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						

Table 6.40: Pairwise Wilcoxon T-test for *Strasse* and *Temporal Distance* (complete in table B.8)

significantly from the A3. The significant descriptives in figs. 6.56a and 6.56b show that the A9

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	4.12	7.16	0	0	24	24
A7	302	2.10	5.21	0	0	24	24
A70	50	1.64	5.20	0	0	23	23
A71	10	3.40	4.74	0	0	13	13
A6	198	2.09	5.07	0	0	23	23
A73	86	3.16	6.33	0	0	24	24
A3	1023	1.81	4.92	0	0	24	24
A99	312	2.72	5.91	0	0	24	24
A96	230	3.09	6.17	0	0	24	24
A995	14	1.64	6.15	0	0	23	23
A92	82	1.20	3.85	0	0	22	22
A93	160	2.39	5.50	0	0	23	23
A94	56	2.18	5.06	0	0	23	23

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.56: Group descriptives of *Strasse* and *Temporal Distance*

and A96 have a significantly higher $\bar{x}$ than the A3, A7 and A92. The differences between these groups are only about 2 min. The descriptives also show that the A9, A71, A73 and A96 have considerable higher $\bar{x}$, with a maximum deviation of 2 min from the overall $\bar{x}$. The $\bar{x}$ and σ both show a similar trend, which means that $\bar{x}$ should be representable. In the case of A71 and A92 additional analysis is necessary for reliable results.

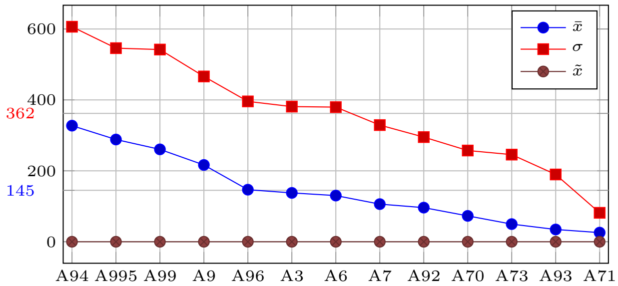
The Kruskal-Wallis test of the relation *Strasse* - *SDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SDist* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.41. It shows that the groups A3, A73, A93, A94 and A99 differ significantly. With the significant descriptives in figs. 6.57a and 6.57b it can be interpreted that the roads A94 and A99 have a

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A73	0.03	1.00	1.00	1.00	0.89											
A99	1.00	0.61	1.00	1.00	1.00	0.03	1.00									
A93	0.00	0.05	1.00	1.00	0.01	1.00	0.00	0.00	0.06	0.16	0.12	1.00				
A94	1.00	0.09	0.72	1.00	0.76	0.00	0.40	1.00	0.30	1.00	1.00	1.00	0.00	1.00		

Table 6.41: Pairwise Wilcoxon T-test for *Strasse* and *Spatial Distance* (complete in table B.9)

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	216.54	466.13	0	0	1988	1988
A7	302	106.12	329.01	0	0	1882	1882
A70	50	72.94	257.26	0	0	1253	1253
A71	10	25.80	81.59	0	0	258	258
A6	198	130.20	379.68	0	0	1968	1968
A73	86	49.74	245.74	0	0	1924	1924
A3	1023	137.75	381.13	0	0	1979	1979
A99	312	260.51	542.02	0	0	1977	1977
A96	230	146.79	395.78	0	0	1958	1958
A995	14	288.29	545.94	0	0	1749	1749
A92	82	96.22	295.14	0	0	1626	1626
A93	160	34.58	189.72	0	0	1769	1769
A94	56	327.38	606.32	0	0	1983	1983

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.57: Group descriptives of *Strasse* and *Spatial Distance*

higher spatial distance (302 m on average) than the A73 and A93. The $\bar{x}$ and σ both show the same trend, which means that $\bar{x}$ should be representable. The descriptives also show that the A94, A995, A99 and A9 have considerable higher and the A70, A73, A93, A71 considerable lower $\bar{x}$ when compared to the overall $\bar{x}$ of 145 m.

The Kruskal-Wallis test of the relation *Strasse* - *Cov* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *Cov* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.42. It shows

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A70	0.01	1.00														
A3	0.00	0.00	0.00	1.00	0.01	1.00										
A99	0.00	0.00	0.00	0.15	0.00	0.00	0.00									
A96	0.00	0.18	1.00	1.00	0.00	0.00	0.00	0.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.02	1.00							
A92	0.00	1.00	1.00	1.00	0.04	0.05	0.00	0.00	1.00	1.00						
A93	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	1.00				
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.36	1.00	1.00	1.00	0.00	1.00		

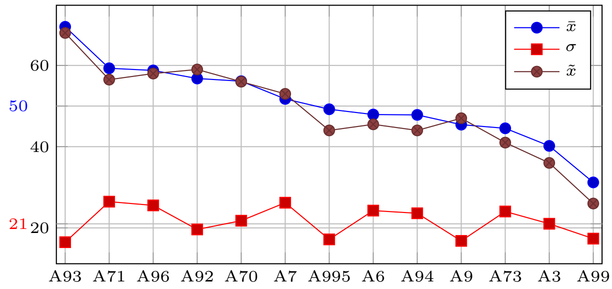
Table 6.42: Pairwise Wilcoxon T-test for *Strasse* and *Coverage* (complete in table B.10)

that the groups A3, A7, A70, A73, A9, A92, A93, A94, A96, A99 and A995 differ significantly. With the descriptives in figs. 6.58a and 6.58b it can be interpreted that the road A70, A92, A93, and A96 have a significantly higher coverage than the A3, A73 and A99.

The Kruskal-Wallis test of the relation *Strasse* - *TLCar* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TLCar* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.43. It shows that the groups A3, A7, A70, A73, A92, A93, A94, A96 and A99 differ significantly. Although the groups differ significantly, the descriptives in figs. 6.59a and 6.59b don't provide

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	45.42	16.80	47.00	7	100	93
A7	302	51.75	26.20	53.00	6	100	94
A70	50	56.12	21.77	56.00	15	100	85
A71	10	59.30	26.46	56.50	18	100	82
A6	198	47.91	24.26	45.50	9	100	91
A73	86	44.52	24.04	41.00	7	99	92
A3	1023	40.19	21.02	36.00	4	100	96
A99	312	31.20	17.36	26.00	4	93	89
A96	230	58.77	25.54	58.00	7	100	93
A995	14	49.21	17.17	44.00	26	73	47
A92	82	56.78	19.62	59.00	12	100	88
A93	160	69.51	16.48	68.00	22	91	69
A94	56	47.82	23.60	44.00	10	100	90

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

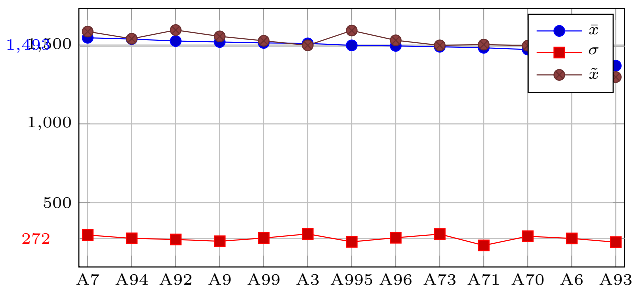
Figure 6.58: Group descriptives of *Strasse* and *Coverage*

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A93	0.00	0.00	1.00	1.00	0.10	0.52	0.00	0.00	0.00	1.00	0.00	1.00				
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00		

Table 6.43: Pairwise Wilcoxon T-test for *Strasse* and *Time-loss Car* (complete in table B.11)

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	1519.32	255.65	1555.00	1001	1994	993
A7	302	1545.96	295.36	1585.50	1000	1999	999
A70	50	1471.40	287.41	1496.00	1015	1984	969
A71	10	1482.50	229.21	1503.00	1132	1819	687
A6	198	1459.12	273.55	1438.50	1013	1997	984
A73	86	1489.05	300.96	1498.00	1003	1974	971
A3	1023	1510.49	302.14	1498.00	1000	1999	999
A99	312	1513.44	276.21	1527.00	1005	1991	986
A96	230	1494.62	278.01	1530.00	1010	1997	987
A995	14	1497.86	252.08	1592.00	1034	1813	779
A92	82	1525.43	267.00	1595.00	1056	1976	920
A93	160	1368.61	249.94	1297.00	1012	1988	976
A94	56	1537.91	274.00	1540.00	1079	1990	911

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.59: Group descriptives of *Strasse* and *Time-loss Car*

interpretable differences, because all groups descriptives are very similar. The descriptives values are also very similar to the $\bar{x}$, σ or $\tilde{x}$ respectively.

The Kruskal-Wallis test of the relation *Strasse* - *TLHGV* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TLHGV* between the groups of *Strasse*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.44. It shows that the groups A3, A7, A70, A73, A92, A93, A94, A96 and A99 differ

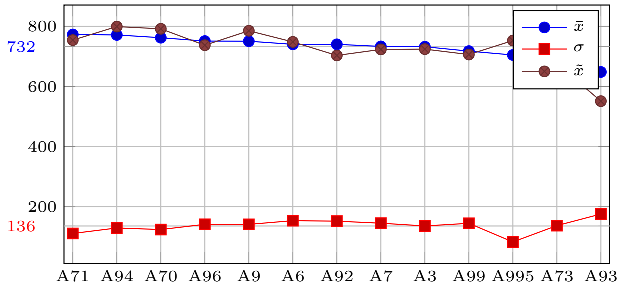
	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A93	0.00	0.00	0.00	0.62	0.00	0.01	0.00	0.00	0.00	1.00	0.00	1.00				
A94	1.00	1.00	1.00	1.00	1.00	0.57	1.00	0.79	1.00	1.00	1.00	1.00	0.00	1.00		

Table 6.44: Pairwise Wilcoxon T-test for *Strasse* and *Time-loss HGV* (complete in table B.12)

significantly. Although the groups differ significantly, the descriptives in figs. 6.60a and 6.60b don't provide interpretable differences, because all groups descriptives are very similar. The descriptives values are also very similar to the $\bar{x}$, σ or $\tilde{x}$ respectively.

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
A9	656	750.33	141.34	785.00	506	997	491
A7	302	732.96	145.27	723.00	500	997	497
A70	50	761.96	124.17	791.50	540	955	415
A71	10	772.70	111.11	754.00	565	981	416
A6	198	739.95	153.70	748.00	504	996	492
A73	86	702.86	137.21	676.00	502	964	462
A3	1023	731.93	136.09	724.00	501	996	495
A99	312	717.33	144.89	706.00	501	998	497
A96	230	750.74	141.48	737.00	500	999	499
A995	14	704.50	82.95	752.50	567	810	243
A92	82	739.94	152.04	703.00	529	979	450
A93	160	648.00	175.73	551.00	500	999	499
A94	56	771.16	129.38	799.00	559	959	400

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.60: Group descriptives of *Strasse* and *Time-loss Heavy goods vehicle*

Month

This section analyzes the correlated relations of the accident variable *Month*. The Kruskal-Wallis test of *Month* - *TDist* results in a p-value above the defined α-level. Therefore the null hypothesis can't be rejected and no significant differences in *TDist* for the relations of *Month* are present.

The Kruskal-Wallis test of the relation *Month* - *TAvg* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TAvg* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.45. It shows that all groups have significantly differences form at least on other group. The month of

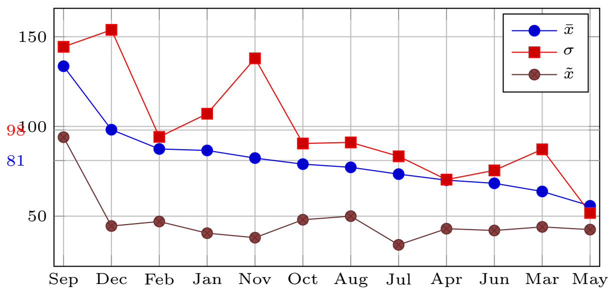
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Sep	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
Oct	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00		
Nov	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	
Dec	1.00	1.00	0.97	1.00	0.01	1.00	0.20	1.00	0.00	1.00	1.00

Table 6.45: Pairwise Wilcoxon T-test for *Month* and *Average Temporal Extent* (complete in table B.13)

September has significantly differences to almost all other groups. This means that all months are significantly related and September is very significantly related to the average temporal extend. The significant descriptives in figs. 6.61a and 6.61b show that the $\bar{x}$ of September deviated heavily

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	118	86.61	107.08	40.50	6	630	624
Feb	160	87.44	94.17	47.00	4	394	390
Mar	216	63.76	87.26	44.00	6	673	667
Apr	271	70.06	70.37	43.00	7	381	374
May	268	55.76	51.77	42.50	4	399	395
Jun	186	68.30	75.52	42.00	3	515	512
Jul	529	73.41	83.37	34.00	5	575	570
Aug	273	77.21	91.11	50.00	4	581	577
Sep	370	133.56	144.37	94.00	3	966	963
Oct	307	78.98	90.51	48.00	5	447	442
Nov	213	82.36	137.96	38.00	5	1326	1321
Dec	278	98.18	153.84	44.50	5	1326	1321

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.61: Group descriptives of *Month* and *Average Temporal Extent*

from the mean, with 52 min. Therefore it can be interpreted that roadworks in September result in significantly longer jam durations than the average. The descriptives also show that April, May, June and July have considerable lower $\bar{x}$ than the average. Therefore it can be interpreted that roadworks in April, May, June and July result be significantly shorter jam durations than the average.

The Kruskal-Wallis test of the relation *Month* - *SMax* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SMax* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.46. It shows

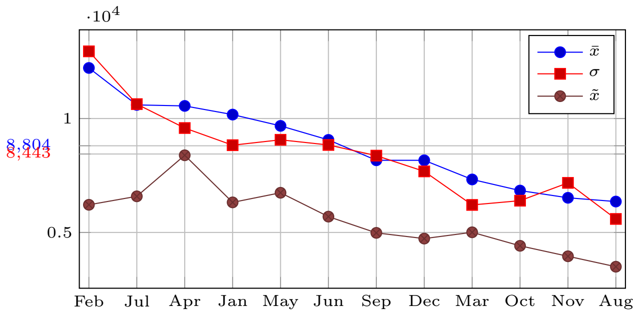
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Apr	1.00	1.00	0.01								
Aug	0.00	0.00	0.69	0.00	0.00	1.00	0.00				
Sep	0.37	0.01	1.00	0.00	0.07	1.00	0.00	1.00			
Oct	0.02	0.00	1.00	0.00	0.01	1.00	0.00	1.00	1.00		
Nov	0.00	0.00	0.02	0.00	0.00	0.01	0.00	1.00	0.14	1.00	
Dec	0.68	0.30	1.00	0.01	1.00	1.00	0.01	1.00	1.00	1.00	0.05

Table 6.46: Pairwise Wilcoxon T-test for *Month* and *Maximal Spatial Extent* (complete in table B.14)

that all groups have significant differences form some other group. The groups of February, April, July, August, October and November differ from a relative high number of significant groups and are therefore strongly significant related to the maximal spatial extend. With the

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	118	10172.38	8829.63	6320.00	1109	44751	43642
Feb	160	12219.37	12948.08	6214.00	999	48987	47988
Mar	216	7325.81	6207.77	5007.00	1436	31160	29724
Apr	271	10553.42	9582.49	8387.00	1350	47196	45846
May	268	9677.59	9065.93	6742.00	1035	40805	39770
Jun	186	9062.72	8839.51	5695.00	1084	34411	33327
Jul	529	10596.79	10632.15	6586.00	902	49765	48863
Aug	273	6356.36	5593.06	3497.00	699	32154	31455
Sep	370	8167.19	8370.94	4982.00	965	41011	40046
Oct	307	6838.37	6394.64	4414.00	1025	31676	30651
Nov	213	6521.21	7176.02	3955.00	1014	33764	32750
Dec	278	8165.85	7680.25	4731.00	951	34227	33276

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.62: Group descriptives of *Month* and *Maximal Spatial Extent*

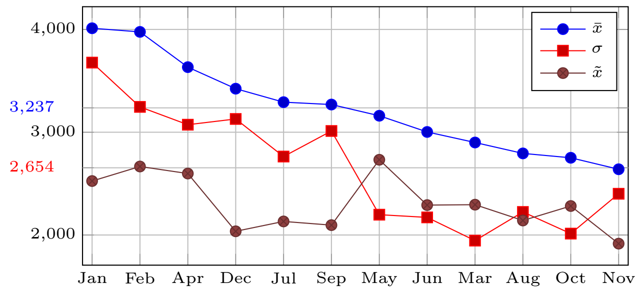
descriptives in figs. 6.62a and 6.62b it can be interpreted that roadworks in February, April and July result in significantly spatial longer jams than the average. The roadwork jams in March, August, October and November tend to be significantly spatial shorter than the average. The Kruskal-Wallis test of the relation *Month* - *SAvg* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SAvg* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.47. It shows that most groups have significantly differences to another group. The groups of February, April and November differ from a relative high number of groups. With the descriptives in figs. 6.63a and 6.63b it can be interpreted that roadworks in January, February and April result in significantly spatial longer jams than the average. The roadwork jams in March, August, October and November tend to be significantly spatial shorter than the average. When comparing the

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Aug	0.25	0.00	1.00	0.00	1.00	1.00	0.51				
Oct	1.00	0.03	1.00	0.03	1.00	1.00	1.00	1.00	1.00		
Nov	0.00	0.00	0.02	0.00	0.01	0.15	0.00	1.00	0.01	0.09	

Table 6.47: Pairwise Wilcoxon T-test for *Month* and *Average Spatial Extent* (complete in table B.15)

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	118	4011.67	3677.51	2525.00	651	14785	14134
Feb	160	3976.34	3246.48	2666.50	718	14150	13432
Mar	216	2899.98	1945.37	2295.00	676	9748	9072
Apr	271	3633.27	3073.18	2598.00	569	15602	15033
May	268	3160.32	2197.68	2732.00	458	10860	10402
Jun	186	3002.61	2171.31	2291.00	692	11206	10514
Jul	529	3292.55	2763.41	2132.00	502	13146	12644
Aug	273	2793.92	2224.25	2141.00	506	12288	11782
Sep	370	3269.36	3012.09	2096.00	404	14776	14372
Oct	307	2751.38	2013.59	2282.00	625	11161	10536
Nov	213	2639.70	2401.25	1916.00	305	13639	13334
Dec	278	3423.72	3129.03	2036.00	284	15054	14770

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.63: Group descriptives of *Month* and *Average Spatial Extent*

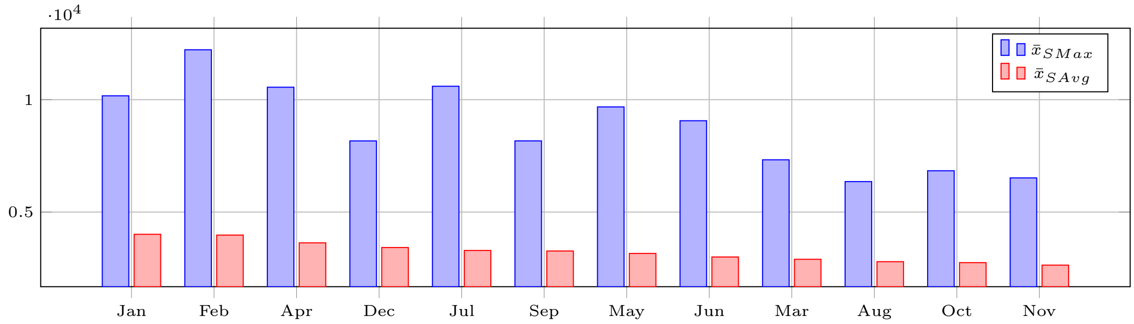


Figure 6.64: Comparison of descriptives $\bar{x}_{SMax}$ and $\bar{x}_{SAvg}$ (*SMax/SAvg* by *Month*)

$\bar{x}$ values of the maximal and average (spatial) extend (shown in fig. 7.10) it becomes clear that the average variable has considerable lower values than the maximum variable, which is to be expected. The distribution shows a that there are considerable difference in the jam maximal lengths, but not as much in average lengths. There is a distinctive trend of magnitude in the maximal variable, but not in the same order as in the average.

The Kruskal-Wallis test of the relation *Month* - *SDist* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *SDist* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.48. It shows

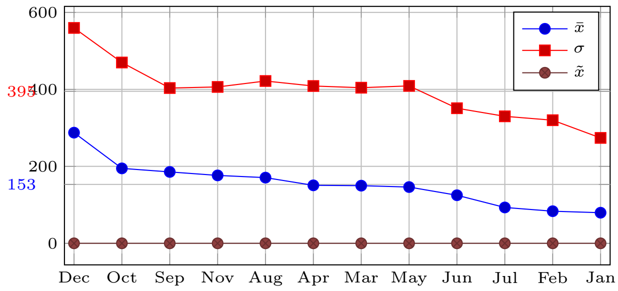
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Jul	1.00	1.00	0.18	0.12	0.02	1.00					
Aug	0.36	0.01	1.00	1.00	1.00	1.00	0.00				
Sep	0.21	0.00	1.00	1.00	1.00	1.00	0.00	1.00			
Oct	1.00	0.05	1.00	1.00	1.00	1.00	0.00	1.00	1.00		
Nov	0.00	0.00	0.03	0.01	0.03	0.01	0.00	0.68	0.73	0.09	
Dec	0.01	0.00	0.20	0.10	0.24	0.04	0.00	1.00	1.00	0.68	1.00

Table 6.48: Pairwise Wilcoxon T-test for *Month* and *Spatial Distance* (complete in table B.16)

that most groups have significantly differences to other groups. The groups of February, July, November and December differ from a relative high number of groups. With the descriptives in

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	118	79.52	273.88	0	0	1686	1686
Feb	160	83.44	319.89	0	0	1615	1615
Mar	216	149.79	404.38	0	0	1861	1861
Apr	271	150.68	408.76	0	0	1910	1910
May	268	146.19	409.04	0	0	1979	1979
Jun	186	124.91	351.03	0	0	1879	1879
Jul	529	93.17	330.01	0	0	1973	1973
Aug	273	170.69	421.76	0	0	1859	1859
Sep	370	185.51	403.48	0	0	1968	1968
Oct	307	194.72	469.83	0	0	1983	1983
Nov	213	176.46	406.35	0	0	1925	1925
Dec	278	287.55	559.68	0	0	1988	1988

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.65: Group descriptives of *Month* and *Spatial Distance*

figs. 6.65a and 6.65b it can be interpreted that roadworks in October and December result in significantly longer spatial distances than the average. The roadwork jams in January, February and July tend to have significantly shorter spatial distances than the average.

The Kruskal-Wallis test of the relation *Month* - *Cov* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *Cov* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.49. It

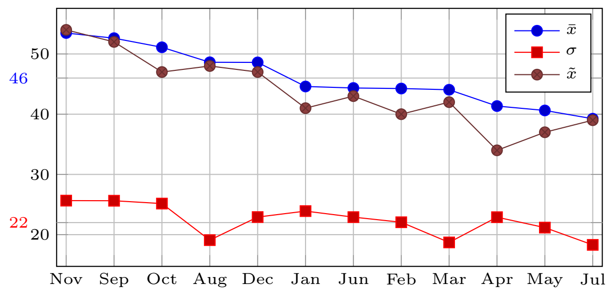
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Aug	0.42	0.32	0.11	0.00	0.00	0.35	0.00				
Sep	0.14	0.03	0.03	0.00	0.00	0.02	0.00	1.00			
Oct	0.50	0.35	0.20	0.00	0.00	0.14	0.00	1.00	1.00		
Nov	0.06	0.01	0.00	0.00	0.00	0.03	0.00	0.35	1.00	1.00	
Dec	1.00	1.00	1.00	0.01	0.00	1.00	0.00	1.00	1.00	1.00	0.79

Table 6.49: Pairwise Wilcoxon T-test for *Month* and *Coverage* (complete in table B.17)

shows that most groups have significantly differences to other groups. The groups of April, May, July, September, November and December differ from a relative high number of groups. Only January doesn't show any significant differences. With the descriptives in figs. 6.66a and 6.66b

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	118	44.60	23.91	41	7	100	93
Feb	160	44.26	22.06	40	9	100	91
Mar	216	44.06	18.71	42	8	100	92
Apr	271	41.35	22.90	34	8	100	92
May	268	40.62	21.17	37	7	100	93
Jun	186	44.35	22.91	43	4	100	96
Jul	529	39.25	18.30	39	4	100	96
Aug	273	48.62	19.09	48	7	100	93
Sep	370	52.62	25.63	52	9	100	91
Oct	307	51.11	25.16	47	8	100	92
Nov	213	53.48	25.65	54	6	100	94
Dec	278	48.60	22.92	47	7	100	93

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.66: Group descriptives of *Month* and *Coverage*

it can be interpreted that roadworks in September, October, November and December result in significantly denser jams than the average. The roadwork jams in March and July tend to be significantly less dense jams than the average.

The Kruskal-Wallis test of the relation *Month* - *TLCar* produces a p-value below .0001, which is far below the defined α-level. Therefore the null hypothesis can be rejected, meaning that there are strong significant differences in the variable *TLCar* between the groups of *Month*. The significant groups can be identified with a pairwise Wilcoxon T-test, which is shown in table 6.50. It shows that most groups have significantly differences to other groups. The groups of February,

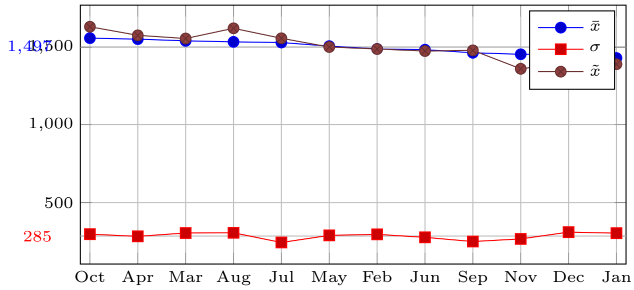
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Mar	0.05	1.00									
Apr	0.01	1.00	1.00								
Jul	0.05	1.00	1.00	1.00	1.00	1.00					
Sep	1.00	1.00	0.08	0.00	1.00	1.00	0.00	0.04			
Oct	0.04	1.00	1.00	1.00	1.00	0.40	0.73	1.00	0.00		
Nov	1.00	1.00	0.10	0.03	1.00	1.00	0.00	0.16	1.00	0.00	
Dec	1.00	1.00	0.08	0.00	1.00	1.00	0.04	0.28	1.00	0.02	1.00

Table 6.50: Pairwise Wilcoxon T-test for *Month* and *Time-loss Car* (complete in table B.18)

May and June don't show any significant differences. The descriptives in figs. 6.67a and 6.67b

Group	n	$\bar{x}$	σ	$\tilde{x}$	min	max	Δ
Jan	118	1428.22	303.79	1388.50	1012	1966	954
Feb	160	1486.67	295.94	1486.00	1005	1997	992
Mar	216	1538.59	304.77	1553.50	1010	1997	987
Apr	271	1549.43	283.26	1574.00	1001	1976	975
May	268	1504.66	289.40	1499.00	1003	1999	996
Jun	186	1481.70	276.89	1473.00	1021	1994	973
Jul	529	1527.63	243.81	1555.00	1001	1999	998
Aug	273	1531.97	306.05	1619.00	1000	1989	989
Sep	370	1461.87	249.90	1477.50	1001	1982	981
Oct	307	1555.69	297.27	1629.00	1007	1996	989
Nov	213	1452.33	266.82	1359.00	1036	1988	952
Dec	278	1452.09	310.35	1399.00	1000	1997	997

(a) Table of all descriptives



(b) Plot of descriptives $\bar{x}$, σ and $\tilde{x}$

Figure 6.67: Group descriptives of *Month* and *Time-loss Car*

show just small differences. Although the groups show significant differences the effects in the descriptives are too small for a meaningful interpretation.

6.2.6 Congestion - Roadworks categorizes as Jam Initiator

In the methodology section 5.3 it was stated that a classification of congestion - roadwork matched would be viable to analysis congestion which are specifically allocated spatial and temporal in from of a roadwork. Unfortunately the timeframe of the these didn't allow for the integrating the results. The analysis was finished just in time of finishing the thesis, however this did not leave time for implementing the analysis results in the textual part of the thesis, yet the results of the analysis are available in the repository linked in the introduction. A quick review of the *Jam Initiator* analysis results did not show any significant differences of the matched dataset results, which mitigates the scientific damage of skipping this planned part of the analysis.

7 Summary of analysis

In the previous chapter the datasets where analyzes for collections and the relevance of the correlation was evaluated in detail. The analysis has presented are a number of separated statistical statements and interpretations, which need to be summarizes to comprise a general statement. This summary and general interpretation is topic of this chapter. Since this chapter is a summary of interpretations and is is based on the analysis in the previous chapter only significant correlations which hold a relevant interpretability are summarized and not every analysis detail is repeated.

7.1 Accidents

The analysis of the congestion - accidents correlations is separated into sections, to account for the different accident types of Initiator, Effector and Follower (see section 5.3). In this section, the findings for each accident characteristic from the different categories are collected, summarizes and interpreted.

7.1.1 Street (*Strasse*)

The relation classes of Jam Intiator and Jam Follower don't show any relevance in the found collections. In the case of the Jam Intiator this it especially unfortunate, since it is the most promising group for predictions. The relations in the complete matched dataset and the relation class of Jam Effector are summarizes in the following.

Duration : The maximum and average jam duration is generally related to the street, but not all streets show statistical significant differences. Therefore there is no general significant trend to be interpreted but since the relation has a general significance and individual significant differences are present it can be assumed, that the descriptives are generally representable. When comparing all descriptive means (shown in fig. 7.1) it becomes clear, that the maximum and average variables differ by high ranges, which is to be expected, but also show a similar tend. The comparison of the global and Jam Effectors variables presents an general increase of means from the global to Jam Effectors variable. This means that accidents during a congestion are associated with (temporal) longer jams, than jams in general.

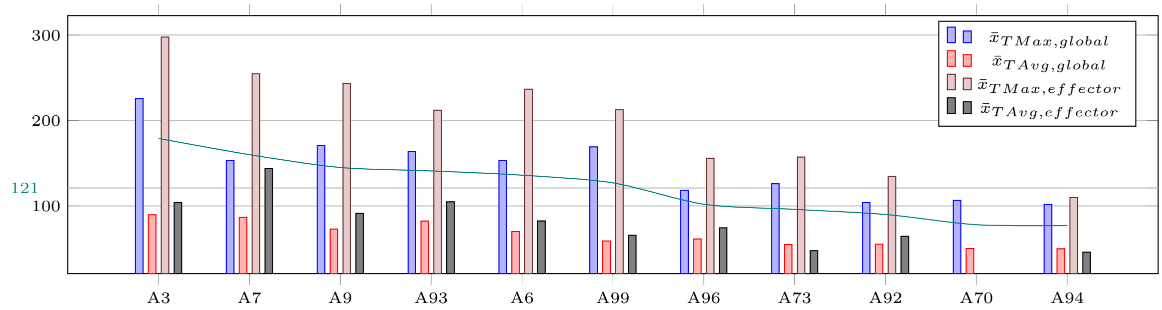


Figure 7.1: Comparison of descriptives $\bar{x}_{TMax,global}$, $\bar{x}_{TAvg,global}$, $\bar{x}_{TMax,effector}$ and $\bar{x}_{TAvg,effector}$ by *Strasse*

Length : The maximum and average jam length is generally related to the street, but not all streets show statistical significant differences. Therefore there is no general significant trend to be interpreted but since the relation has a general significance and individual significant differences are present it can be assumed, that the descriptives are generally representable. When comparing all descriptive means (shown in fig. 7.2) it becomes clear, that the maximum and average variables differ by high ranges, which is to be expected, but also show a similar tend. The comparison of

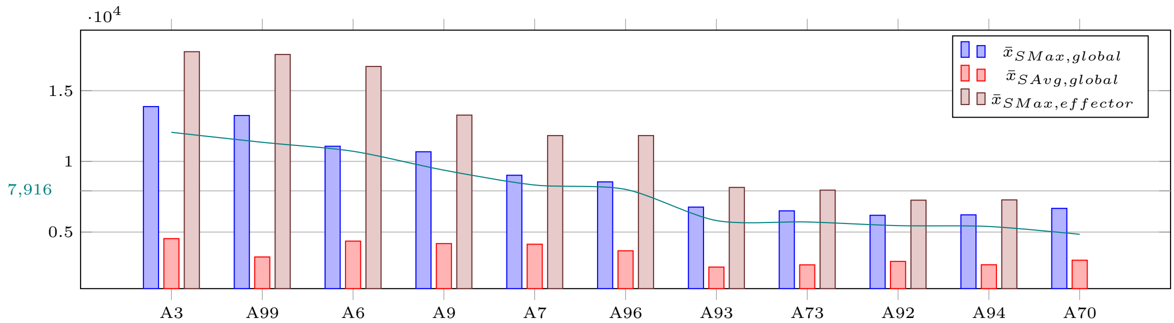


Figure 7.2: Comparison of descriptives $\bar{x}_{SMax,global}$, $\bar{x}_{SAvg,global}$ and $\bar{x}_{SMax,effector}$ by *Strasse*

the global and Jam Effectors variables presents like with the jam duration an general increase of means from the global to Jam Effectors variable. This means that accidents during a congestion are associated with (spatial) longer jams, than jams in general.

Coverage : The jam coverage is generally related to the street, with most road showing statistical significant differences. Therefore it can be assumed, that the descriptives are generally representable. The comparison of the global and Jam Effectors variables (shown in fig. 7.3) presents an general decrease of means from the global to Jam Effectors variable. This means that accidents during a congestion are associated with less denser jams, than jams in general. In accordance with these finding, it can be interpreted that in general the street plays a significant role defining the size and form of a congestion. This is not surprising since the demand on the street, which varies heavily between the streets is the major factor influencing the buildup of a congestion. Non the less it is interesting the the differences are substantial with ranges up to 60 % increases between groups and are replicated over difference variable relations. Another effect it the increased jam duration and length with the Jam Effectors in comparison with jams in general. This is due to either the higher probability of a accident in very long jams or that

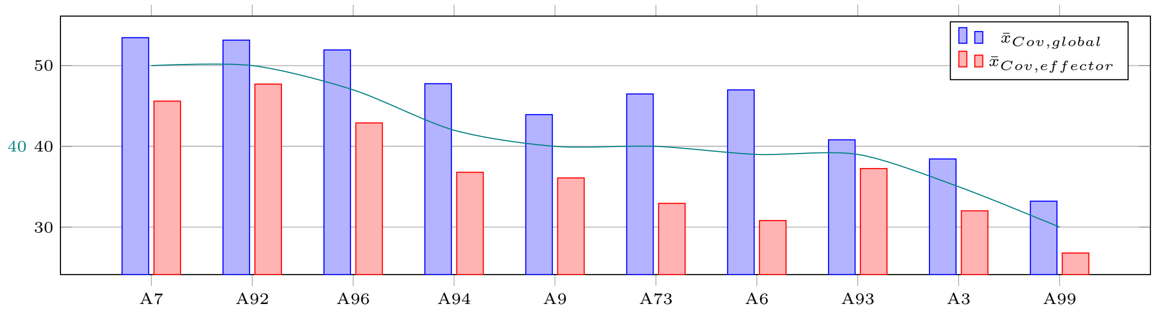


Figure 7.3: Comparison of descriptives $\bar{x}_{Cov,global}$ and $\bar{x}_{Cov,effector}$ by *Strasse*

accident during a can jam increase the jam extends itself.

7.1.2 Accident category (*Kat*)

The variable *Kat* showed a general dependence to the maximum and average duration and a general trend of increasing duration with the injury gravity of the accident in the global and *Jam Initiators* dataset. In the global dataset the category of accidents with property damage does not fit into this trend and sits in-between the lightly and heavily injured category. This changes when considering just the *Jam Initiators*, where all categories follow one increasing trend. When comparing all descriptive means (shown in fig. 7.4) it becomes clear, that the maximum and average variables differ by high ranges, which is to be expected. The comparison also show that that the jam duration significant increases with the gravity of the accident. The category of deadly accident differs by 130 min from accidents of lightly injured and property damage accidents. Unlike the temporal length, the spatial length does not correlate with the

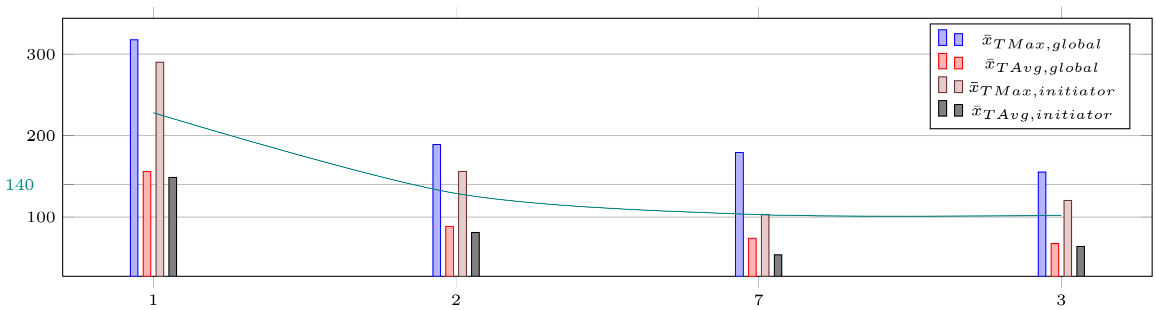


Figure 7.4: Comparison of descriptives $\bar{x}_{TMax,global}$, $\bar{x}_{TAvg,global}$, $\bar{x}_{TMax,initiator}$ and $\bar{x}_{TAvg,initiator}$ by *Kat*

accident category. According to the global dataset, the temporal distance is heavily related to the accident category, but the descriptives only show a maximum differences of 6 min, which is arguably low for an interpretation. Non the less in can be stated, that the temporal distance between accident and congestion increased significantly with the injury gravity. The variable also correlates in other datasets, but without significant differences.

7.1.3 Accident type (*Typ*)

The *Typ* only significantly relates to the temporal distance between the accident and congestion. In the global dataset accidents of the kind *Driving accident* and *Other* (average of 8.5 min) have a longer temporal distance than crossing accident or accidents in straight traffic (average of 3.5 min). A significant differences of similar kinds can be also observed in the Jam Intiator dataset with a temporal distance of *driving* and *other* accidents (average of 12 min) compared to *merging*, *crossing* and *straight traffic* accidents (average of 8 min). The variable also correlates with other variables, but without significant differences. When comparing all descriptive means (shown in fig. 7.5) it becomes clear, that the temporal distance increases for *Jam Initiators* and the trend shown in both dataset is mirrored.

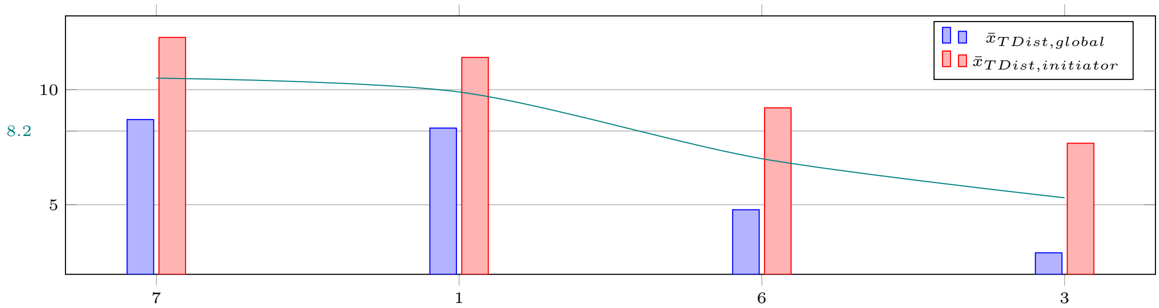


Figure 7.5: Comparison of descriptives $\bar{x}_{TDist,global}$ and $\bar{x}_{TDist,initiator}$ by *Typ*

7.1.4 Accident kind (*UArt*)

The global dataset shows correlations of the accident kind and the temporal distance as well as the coverage, which are both significant. Collisions with *starting*, *standing*, *stopping*, *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles take an average of 5 min to form jams, when accident collisions with *obstacles* or *left/right* nearby vehicles take an average of 8 min to form a congestion. Accidents of the category *turning* and *crossing* vehicles have the most immediate reaction of 2.5 min. The *Jam Initiator* dataset show the same features of accident collisions with *starting*, *standing*, *stopping*, *turning*, *crossing*, *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles having an average of 9 min and accident collisions with *obstacles* or *left/right* nearby vehicles an average of 12 min. This can be also seen in fig. 7.6 where the two dataset variable are compared directly. The coverage relation shows the same grouping like the temporal distance. Accident collisions with *starting*, *standing*, *stopping*, *ahead and waiting vehicle* and *vehicle on separate lane in same direction* vehicles have an average coverage of 41 % and are therefore associated with less dense jams than accident collisions with *obstacles* or *left/right* nearby vehicles, which have a average coverage of 54 %. Jams associated with accidents of the category *turning* and *crossing* vehicles are the least dense with 33 %. The *Jam Initiator* dataset presents the same trend and supports the findings, but without any significant groups. In the *Jam Initiator* dataset *UArt* also correlates significantly with the maximal temporal extend but only in the form that accidents with vehicles on the

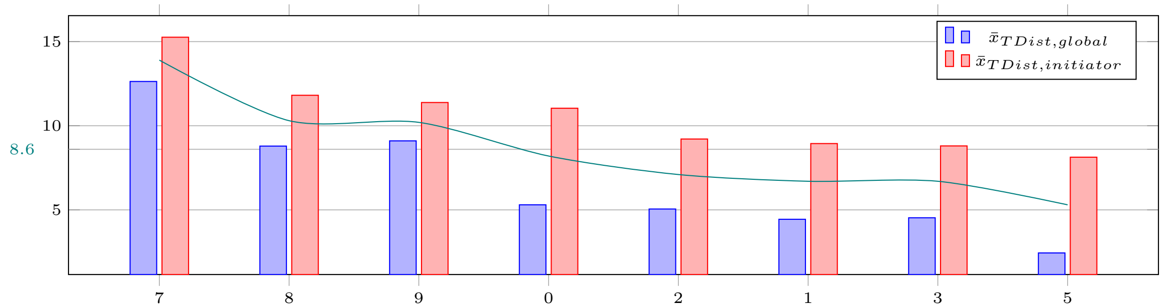


Figure 7.6: Comparison of descriptives $\bar{x}_{TDist,global}$ and $\bar{x}_{TDist,initiator}$ by *UArt*

separate lane in the same direction can be associated with 24 min longer jams, than accidents with *obstacles* in general.

7.1.5 Accident cause (*AUrs*)

The accident cause describing variable *AUrs* is related to the average spatial extend, temporal distance and coverage, but without significant group differences. The descriptives show minor features like that the jam coverage decreases with the accident causes *Slippery street due to rain* to *Cart track due to rain, snow or ice* to by 13 % on average in the *Jam Initiator* dataset. The global dataset show the same feature with an average decrease of 17 %. The global dataset they also reveals that the (spatial) jam length increases with the accident causes *Slippery street due to rain* over *Visibility issues due to sun or glare* to *Slippery street due to snow or ice* by 2382 m on average. The temporal distance between accident and congestion decreases with the accident causes *Slippery street due to snow or ice* and *Slippery street due to rain* (average of 7.7 min) to *Visibility issues due to sun or glare* to by 3 min on average.

7.1.6 Accident collision object type (*AufHi*)

The variable *AufHi* describes the type of collision object and is related to the temporal distance and coverage in the global dataset. The temporal distance between jams and accidents is 4.7 min for *trees* and 8.4 min for *guardrail* or *other obstacles*. The jam coverage of collisions with *trees* (40 % coverage) is 10 % lower than the 50 % coverage of the groups of *guardrail* or *other obstacles*. The variable also correlates with other variables, but without meaningful interpretation or significance.

7.1.7 Accident environment characteristic (*Char*)

The accident characteristic variable correlates significantly with the temporal distance in the *Jam Initiator* dataset, but without groups specific difference.

7.1.8 Accident lighting environment (*Lich*)

The lighting situation correlates significantly with the coverage and shows that jams in darkness are 18 % denser than jams in daylight. The state of the street lighting show a similar effect of 10 % denser jams in case of broken street lighting.

7.1.9 Road condition (*Zust*)

The road condition correlates with the coverage of a jam, in the way that the coverage increases by 20 %-25 % from *dry* over *wet* to *ice*.

7.1.10 Weekday (*WoTag*)

The week day correlates with the spatial and temporal extends as well as the temporal distance, coverage and time loss in multiple datasets. But only the the coverage and time loss (HGV) correlation show significant differences. The coverage of jams on Monday's, Saturday's and Sunday's is 7 %-8 % higher than on other week days. The variable also correlates with other variables, but without meaningful interpretation or significance.

7.1.11 Month (*Month*)

The month of the accident correlates in multiple datasets and with most congestion variables, but only in the *Jam Effector* the maximal and average spatial extend show significant differences. The months of January and November have the shortest extends when July and December have the longest. Together with the other months they do not form a distinctive trend.

7.2 Roadworks

Unlike the analysis of the congestion - accident correlations the roadwork correlations are not separated into global and *Jam Initiator*. Although the methodology section 5.3 stated that a classification of congestion - roadwork matched would be viable to analysis congestion which are specifically allocated spatial and temporal in from of a roadwork, the timeframe of the these didn't allow for the integrating the results. The analysis was finished just in time of finishing the thesis, however this did not leave time for implementing the analysis results in the textual part of the thesis, yet the results of the analysis are available in the repository linked in the introduction. Also, as stated in section 6.2.6 a quick review of the *Jam Initiator* analysis did not show breakthought results and is therefore skipped in chapter 6 and this chapter.

7.2.1 Street (*Strasse*)

The road of the roadwork correlates strongly and significantly with all of the congestion characteristics, due to the very high number of samples.

Duration : The maximum and average jam duration is generally related to the street of the roadwork, but not all months show statistical significant differences. Therefore there is no general significant trend to be interpreted but since the relation has a general significance and individual significant differences are present it can be assumed, that the descriptives are generally representable. When comparing all descriptive means (shown in fig. 7.7) it becomes clear, that the maximum and average variables differ by high ranges, which is to be expected, but also show a similar tends. The diagram is sorted by the mean of maximal and average duration and there-

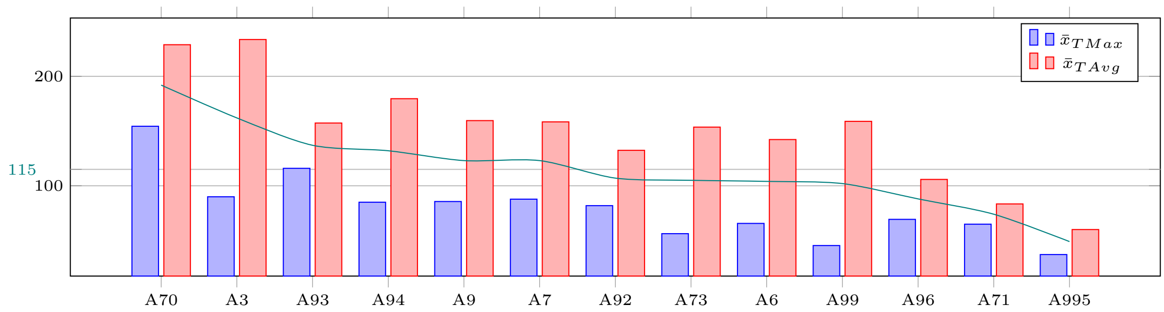


Figure 7.7: Comparison of descriptives $\bar{x}_{TMax}$ and $\bar{x}_{TAvg}$ by *Strasse*

fore shows that the A70 and A3 have considerable longer jams and the A96, A71 and A995 considerable shorter jams in comparison to the mean.

Length : The maximum and average jam length is generally related to the street of the roadwork, but not all streets show statistical significant differences. Therefore there is no general significant trend to be interpreted but since the relation has a general significance and individual significant differences are present it can be assumed, that the descriptives are generally representable. When comparing all descriptive means (shown in fig. 7.8) it becomes clear, that the maximum and average variables differ by high ranges, which is to be expected, but also show a similar tends. The diagram is sort by the mean of maximal and average length like with the

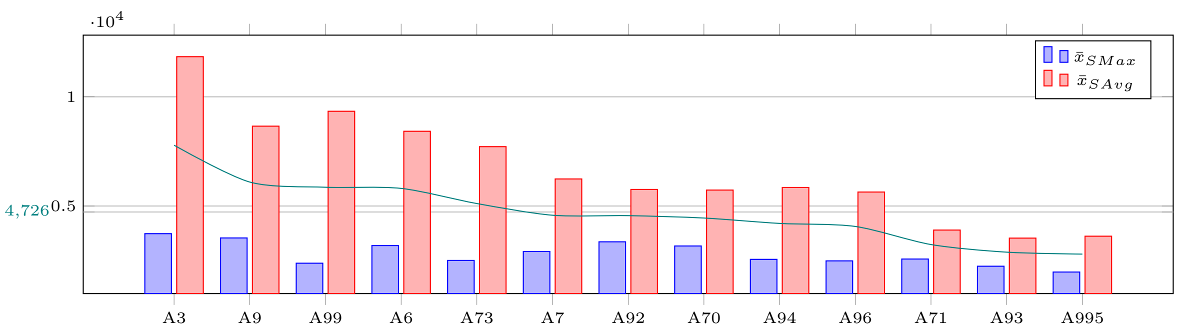


Figure 7.8: Comparison of descriptives $\bar{x}_{SMax}$ and $\bar{x}_{SAvg}$ by *Strasse*

duration diagram and shows that the A3, A9, A99 and A6 have considerable longer jams in comparison to the mean. The A71, A93 and A995 have considerable shorter jams in comparison to the mean.

The A9 and A96 have a significantly higher temporal distance than the A3, A7 and A92, but the difference is only about 2 min. The general trend of the temporal distance shows that the A9, A71, A73 and A96 have higher distances, with a maximum deviation of 2 min from the overall $\bar{x}$.

The spatial distance can be interpreted that the roads A94 and A99 have a higher distance (302 m on average) than the A73 and A93. The general trend also that the A94, A995, A99 and A9 have considerable higher and the A70, A73, A93, A71 considerable lower distances when compared to the overall $\bar{x}$ of 145 m.

For the coverage it can be interpreted that the road A70, A92, A93, and A96 have a significantly higher coverage than the A3, A73 and A99. The time-loss of cars and heavy goods vehicles correlate significantly, but don't provide interpretable differences.

7.2.2 Number closed lanes (*AnzGesperrtFs*)

The number of closed lane does not correlate with any congestion characteristic.

7.2.3 Lane diversion (*Einzug*)

The gravity of the lane diversion does not correlate with any congestion characteristic.

7.2.4 Spatial length (*Length*)

The spatial length of the roadwork does not correlate with any congestion characteristic.

7.2.5 Month (*Month*)

The month of the roadwork correlates strongly with most of the congestion characteristics.

Duration : The average jam duration is generally related to the month of the roadwork and almost all months show statistical significant differences. Therefore there is a general significant trend to be interpreted and the descriptives are generally representable. The descriptive means (shown in fig. 7.9) show a general trend. The distribution can be interpreted that the September deviated heavily from the mean, with 52 min. Therefore it can be interpreted that roadworks in September result in significantly longer jam durations than the average. The descriptives also show that April, May, June and July have considerable lower durations than the average. Therefore it can be interpreted that roadworks in April, May, June and July result be significantly shorter jam durations than the average.

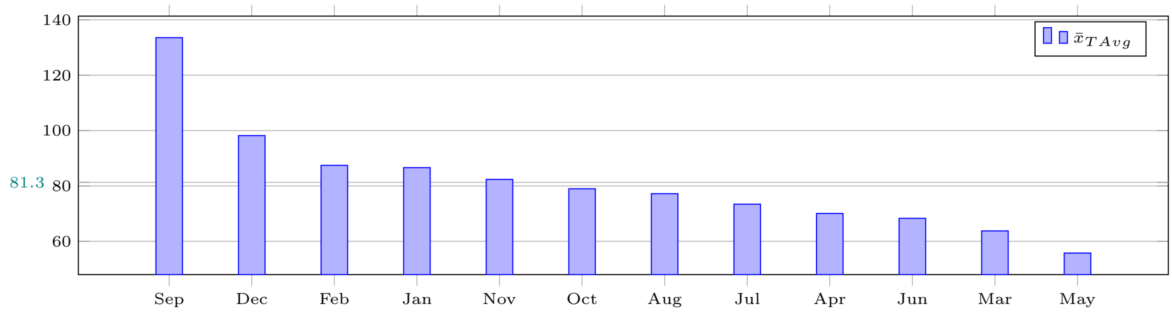


Figure 7.9: Comparison of descriptives $\bar{x}_{TAvg}$ by *Month*

Length : The maximum and average jam length is generally related to the month of the roadwork and almost all months show statistical significant differences. Therefore there is a general significant trend to be interpreted and the descriptives are generally representable. When comparing all descriptive means (shown in fig. 7.10) it becomes clear, that the maximum and average variables differ by high ranges, which is to be expected, but also show a similar tends. The distri-

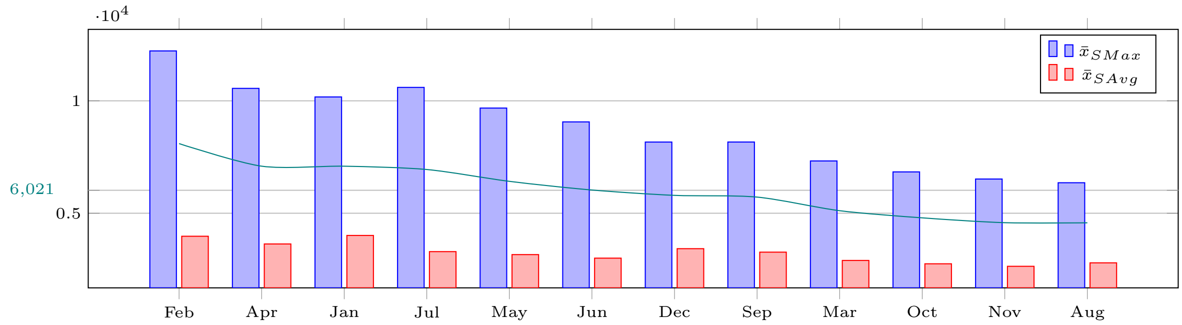


Figure 7.10: Comparison of descriptives $\bar{x}_{SMax}$ and $\bar{x}_{SAvg}$ by *Month*

bution can be interpreted that roadworks in October and December result in significantly longer spatial distances than the average. The roadwork jams in January, February and July result in significantly shorter spatial distances than the average. The coverage of jams shows that roadworks in September, October, November and December result in significantly denser jams than the average. The roadwork jams in March and July tend to be significantly less dense jams than the average. The time-loss correlation shows significance, but very similar descriptives, which means there are no interpretable differences.

7.3 Predictability of accidents and jams

The correlation summary in the sections before showed that there are multiple accident characteristics, which correlate significantly with jams features. Theses significant correlations already provide a statement about the predictions for these relations, but should be further investigated for reliable results. This can be done statistical with hte already introduced Theil's U (see section 3.7). As defined in the statistical theory chapter Theil's U can be only applied on categorical variables which is why the interval characteristics of the congestion object need to be converted into representable categories. This categorization which can also be called binning and is achieved by using the quantile-based discretization function *qcut* from the Pandas framework. The function creates equal sized bins (quantiles) of the measurements, creating an equally good sample sets for all categories.

Based on these categorized measurements (four categories per measurement) from the congestion and the incident characteristics the predictability can be evaluated with Theil's U.

The predictability processing which is similar to the correlation processing, produces a matrix of the predictability U for each relation. In accordance to the defined interpretation of U in section 3.7 any relations with a value close to 1 can be considered as predictable, whereas values closer to 0 are considered to be not predictable. When reviewing the visual representation of the predictability matrixes for accidents (see fig. A.4) and roadworks (see fig. B.4) it becomes clear that none of the relations shows clear predictability, due to a maximum of .34 in all relevant Us. Additionally it should be pointed out that all values but the single .34 are below .2 which very close to 0. The maximum value of .34 is reached by the relation *AUrs2 - TDist*. This means that all relations show no or very low predictability but the relation of *AUrs2 - TDist* which shows at least a low predictability.

8 Conclusion

In conclusion it can be generally stated that there are significant correlations between congestion and incident characteristics, but with many limitations. The summary of the analysis in chapter 7 showed, that general characteristics like Month and Location (Road) are considerably related to congestion characteristics like duration and length. The more specific incident characteristics like accident cause, type and environment or roadwork gravity (number of closed lanes, physical diversion, ...) also often correlate with the congestion characteristics but not significantly in many cases. This missing significance is also present between the categories of the correlated relations/variables.

8.1 Answer of research questions

Correlation In accordance of the summary in chapter 7 and conclusion introduction, the research question of "*Do congestion- and incident-characteristics correlate?*" can be answered with yes, but at the same time it needs to be pointed out that many of the correlations are not significant enough (missing significant differences) for reliable or interpretable results.

The characteristics of Month and Location (Road) however are strongly correlated with the length and duration of jams and also show considerable differences between their groups (e.g. Jan and Feb or A3 and A9). Therefore it can be stated that the significant findings about the correlation of the month and road to the length and duration of jams can be applied on the population (see chapter 7 for detailed findings).

This is not the case for the specific incident characteristics like accident cause, type and environment or roadwork gravity (number of closed lanes, physical diversion, ...). As already mentioned these relation are partly strongly correlated and significant but often do not provide significant differences between their groups. This means that there are significant correlations but without interpretable results. As described in section 3.5.2 this happens when the groups (e.g. $AUrs$ code 1, 2, 3, ...) do not have a sufficient sample size to show the same significance which was found in the general variable ($AUrs$).

This being said there are some incident characteristics which have interpretable significant differences (e.g. *accident lighting environment*, *accident collision object type*, *accident kind*, *accident type* or *accident category*) which are presented in chapter 7.

Prediction The predictability analysis in section 7.3 showed that the characteristics of jams and incidents are not predictable based on the used statistical methods. Therefore the research question of "*Are congestion- and incident-characteristics predictable?*" must be answered with no for this analysis. This being said, the results of the correlation analysis can also provide a statement of predictability with additional research of causality.

8.2 Limitations

Although the methods and tool applied in the evaluation and analysis where developed with the highest standards in mind, it became clear during process that there are a number of limitations to the methodology. As described in the section before the sample size of the datasets limits the quality of the results. The clustering algorithm has some problems with clustering in the spatial dimension and the categorization of the congestion - incident matched is rather simple. All these issues limit the the quality of the result. They also might bias the results but this being said the analysis of the results did not show any indication that one of the apparent issues causes significant effects.

8.3 Improvements

The methods and tools applied in the evaluation and analysis where developed to fit the purpose and timeframe of this thesis which leave much room for improvements. The analysis of the detection algorithm in section 6.1 showed that the calibration is sufficient for a exploratory data analysis approach but did not cover the problems of the detection algorithm which result in wrongly clustered jams. The impact of these were not researched.

Improving the clustering algorithm

The implemented DBSCAN algorithm seems to be suited for provided FCD (see section 6.1) but also show problems with adjacent cluster and overlapping clusters in the spatial dimension. Besides of tuning the calibration and add more rules there is another concept how to improve the results of the algorithm. The current algorithm calculated the distances between point based on the artificial travel time (see section 5.1) which apparently can lead to considerable wrong results because of the usage of the mean speed of all cells instead of the native speed of each cell. This could be solved by using the actual 3D space of time, space and speed for clustering instead of the artificial travel time as 3D like representation.

Another area for improvements is the algorithm performance. An improved version of the DB-SCAN algorithm was present by Yuzhen Zhao, Xiyu Liu and Xiufeng Li[1], which is based on

[1] **https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200751**.

cell-like P systems with promoters and inhibitors. This approach showed a reduction of runtime to $O(n)$ from the implemented $O(n^2)$.

The thesis also focused on the mathematical and statistical methods used for the analysis which resulted in chapter 3. The explained and used methods are definitely suited for the applied analysis but might not be the best option.

Improving the correlation evaluation

Statistical correlation is a broad and heavily discussed field of mathematics. As an result there are often many way and also more complex ways to do statistical analyzes. The intraclass correlation coefficient (ICC) is a quite new way of analyzing the correlation of categorical - categorical relations ships and could be a improved supplement for the correlation coefficient of η and the Kruskal-Wallis h-test[2][3].

The summary of the correlation processing did not only present some significant correlations but also reveals a mayor problem which is the sample size. Although the datasets contain thousands of sample when separating the dataset into specific relation and analysis specific groups the sample size of these groups are often to small for any significant results. This means that for better and reliable results a larger dataset or better distributed dataset is necessary.

8.4 Final thoughts

The thesis showed that congestion and incident characteristics are partly correlated with each other and also show considerable differences. The theory of the statistical methods made clear that proving associations in mixed datasets can be very complex. In contrary to my initial believe these associations where not sufficient for any predictions. In general the thesis showed me that finding and proving associations based on statistical concepts is way harder than expected and the logical believe that incident characteristics must have an impact in jams does not necessary hold up in a statistical analysis. Besides of these learning about data science and statistical methods, the writing of the thesis also showed me with new ways of approaching research and the associated problems.

I want to thank thank my mentor at Schlothauer & Wauer Dipl. - Ing. Stefan Gürtler for helping me with every step of developing the evaluation tool and providing the initial code base. I also want to thank Dipl. - Ing. Johannes Grötsch for providing the necessary datasets and my mentors at the TUM M. Eng. Barbara Karl and Dr. - Ing. Matthias Spangler for assisting me with writing the thesis.

[2] **https://stats.stackexchange.com/questions/73065/correlation-coefficient-between-a-non-dichotomous-nominal-v**
[3] **https://pingouin-stats.org/generated/pingouin.intraclass_corr.html**.

List of Figures

List of Tables

List of Codes

List of Acronyms

ADAC Allgemeine Deutsche Automobil-Club e. V.. 9

ArbIS Arbeitstellenintegrationsystem. 4, 5, 12, 21, 31, 35

ATDM Active Traffic and Demand Management. 1, 2, 5

ATIS Advanced Traveler Information System. 4, 5

BAYSIS Bayerische Straßeninformationssystem. 3, 5, 12, 21, 26, 32, 35

CSV Comma-seperated Values. 31

FCD Floating Car Data. 2, 3, 5, 9, 10, 25, 35, 36, 39

FCO Floating Car Observer. 3

FPD Floating Phone Data. 3

GSM Global System for Mobile Communications. 2

LTE Long Term Evolution. 2

RTTI Real Time Traffic Information. 4

TMC Traffic Messaging Channel. 4

TSM Transportation System Management. 1

UMTS Universal Mobile Telecommunications System. 2

XFCD Extended Floating Car Data. 3

ZVM Zentralstelle Verkehrsmanagement. 3

Bibliography

ADAC (2019). *Staubilanz 2019 – Weniger, dafür längere Staus | ADAC.* <https://www.adac.de/verkehr/verkehrsinformationen/staubilanz/> (visited on 09/07/2020).

Ahlburg, Dennis A. (1984). “Forecast evaluation and improvement using theil’s decomposition”. In: *Journal of Forecasting* 3.3, pp. 345–351. DOI: 10.1002/for.3980030313.

Akoglu, Haldun (09/2018). *User’s guide to correlation coefficients.* DOI: 10.1016/j.tjem.2018.08.001. </pmc/articles/PMC6107969/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107969/>.

ASTRA (2020). *Verkehr - Definitionen.* <https://www.astra.admin.ch/astra/de/home/themen/nationalstrassen/verkehrsfluss-stauaufkommen/definitionen.html> (visited on 06/19/2020).

Baeldung (2020a). *A Guide to TreeMap in Java |.* <https://www.baeldung.com/java-treemap> (visited on 10/30/2020).

— (2020b). *Java TreeMap vs HashMap.* <https://www.baeldung.com/java-treemap-vs-hashmap> (visited on 10/30/2020).

Baguley, Thom and Salvatore S. Mangiafico (2016). *What’s the reason for deriving Cohen’s w from Cramer’s V if the latter is already a measure of effect size?* <https://www.researchgate.net/post/Whats%7B%5C_%7Dthe%7B%5C_%7Dreason%7B%5C_%7Dfor%7B%5C_%7Dderiving%7B%5C_%7DCohens%7B%5C_%7Dw%7B%5C_%7Dfrom%7B%5C_%7DCramers%7B%5C_%7DV%7B%5C_%7Dif%7B%5C_%7Dthe%7B%5C_%7Dlatter%7B%5C_%7Dis%7B%5C_%7Dalready%7B%5C_%7Da%7B%5C_%7Dmeasure%7B%5C_%7Dof%7B%5C_%7Deffect%7B%5C_%7Dsize> (visited on 10/26/2020).

Bardt, Hubertus, Esther Chrischilles, and Manuel Fritsch (2014). *Forschungsberichte aus dem Institut der deutschen Wirtschaft Köln Die Infrastruktur in Deutschland Zwischen Standortvorteil und Investitionsbedarf.* NaN. <http://www.dnb.de>.

Benesty, Jacob et al. (2009). *Pearson Correlation Coefficient.* Springer, Berlin, Heidelberg, pp. 1–4. DOI: 10.1007/978-3-642-00296-0_5. <https://link.springer.com/chapter/10.1007/978-3-642-00296-0%7B%5C_%7D5>.

Benninghaus, Hans (2007). “Die Beschreibung der Beziehung zwischen einer nominalen und einer metrischen Variablen”. In: *Deskriptive Statistik.* VS Verlag für Sozialwissenschaften, pp. 228–249. DOI: 10.1007/978-3-531-90739-0_8. <https://link.springer.com/chapter/10.1007/978-3-531-90739-0%7B%5C_%7D8>.

Bergsma, Wicher (09/2013). "A bias-correction for Cramér's V and Tschuprow's T". In: *Journal of the Korean Statistical Society* 42.3, pp. 323–328. DOI: 10.1016/j.jkss.2012.10.002.

Bliemel, Friedhelm (11/1973). "Theil's Forecast Accuracy Coefficient: A Clarification". In: *Journal of Marketing Research* 10.4, p. 444. DOI: 10.2307/3149394.

BMW (2020). *BMW ConnectedDrive Kundenportal – digitale Vernetzung zu Ihrem BMW.* <https://www.bmw-connecteddrive.de/app/index.html?gclid=Cj0KCQjwoJX8BRCZARIsAEWBFMIylpAiLbXLqB1JRhPbrCtw8dDxctDgnmuKu8NoxCXM1X6VuGC7oDMaAsXxEALw%7B%5C_%7DwcB%7B%5C#%7D/portal/store> (visited on 10/13/2020).

Bortz, Jürgen (2004). *Statistik: für Human- und Sozialwissenschaftler - Jürgen Bortz - Google Books.* <https://books.google.de/books?hl=de%7B%5C&%7Dlr=%7B%5C&%7Did=6s0mBAAAQBAJ%7B%5C&%7Doi=fnd%7B%5C&%7Dpg=PA1%7B%5C&%7Ddq=J%7B%5C%22%7Bu%7D%7Drgen+Bortz:+Statistik+f%7B%5C%22%7Bu%7D%7Dr+Human-+und+Sozialwissenschaftler%7B%5C&%7Dots=a7b9Q8dA-G%7B%5C&%7Dsig=I1HnVe61Jrn2fiTm1VNsXylHJHM%7B%5C#%7Dv=onepage%7B%5C&%7Dq=J%7B%5C%22%7Bu%7D%7Drgen%20Bortz%7B%5C%%7D3A%20Statistik%20f%7B%5C%22%7Bu%7D%7Dr%20Human-%20und%20Sozialwissenschaftler> (visited on 09/17/2020).

Brossart, Daniel F, Vanessa C Laird, and Trey W Armstrong (2017). "Cogent Psychology Interpreting Kendall's Tau and Tau-U for single-case experimental designs". In: DOI: 10.1080/23311908.2018.1518687. <https://www.tandfonline.com/action/journalInformation?journalCode=oaps20>.

Busch, Michael (2005). "Analyse dichtebasierter Clusteralgorithmen am Beispiel von DBSCAN und MajorClust". PhD thesis. Universität Paderborn.

Charlton, Samuel G. and Nicola J. Starkey (07/2013). "Driving on familiar roads: Automaticity and inattention blindness". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 19, pp. 121–133. DOI: 10.1016/j.trf.2013.03.008.

Chauhan, Nagesh Singh (2020). *DBSCAN Clustering Algorithm in Machine Learning.* <https://www.kdnuggets.com/2020/04/dbscan-clustering-algorithm-machine-learning.html> (visited on 10/11/2020).

Cicchetti, Domenic V (1981). *Testing the Normal Approximation and Minimal Sample Size Requirements of Weighted Kappa When the Number of Categories is Large this formula yields the following approximate minimal sample sizes (N) for ordinal scales ranging between 3 and 7 categories: fo.* Tech. rep. 1, pp. 101–104.

Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences Second Edition.* Tech. rep.

Cohen, Jacob et al. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences - Google Books.* <https://books.google.de/books?hl=de%7B%5C&%7Dlr=%7B%5C&%7Did=gkalyqTMXNEC%7B%5C&%7Doi=fnd%7B%5C&%7Dpg=PP1%7B%5C&%7Ddq=Aiken:+Applied+Multiple+Regression+/+Correlation+Analysis+For+The+Behavioral+Scien

ces.%7B%5C&%7Dots=tSBPY2n58g%7B%5C&%7Dsig=-q4R6TR9tnnvlWnJSWwYlqMHzY0%7B%5C#%7Dv=onepage%7B%5C&%7Dq=Aiken%7B%5C%7D3A%20Applied%20Multiple%20Regression> (visited on 09/17/2020).

Dalitz, Christoph and Hochschule Niederrhein (2009). *Kd-Trees for Document Layout Analysis **. Tech. rep. <http://gamera.informatik.hsnr.de/>.

DeJesus, John (2019). *Point Biserial Correlation with Python.* <https://towardsdatascience.com/point-biserial-correlation-with-python-f7cd591bd3b1> (visited on 10/13/2020).

Ellis, Paul D. (2010). *The Essential Guide to Effect Sizes.* Cambridge University Press. DOI: 10.1017/cbo9780511761676. </core/books/essential-guide-to-effect-sizes/72C26CA99366A19CAC4EF5B16AE3297F>.

Erhard, Rainer (2020). *Distanz zwischen zwei Punkten.* <https://sites.google.com/a/erhard-rainer.com/erhard-rainer/mathematik/distanz-zwischen-zwei-punkten> (visited on 10/19/2020).

Farrance, Ian and Robert Frenkel (2012). *Uncertainty of measurement: A review of the rules for calculating Uncertainty components through functional relationships.* </pmc/articles/PMC3387884/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387884/>.

FHA (2011). *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation: Chapter 2.* <https://ops.fhwa.dot.gov/congestion%7B%5C_%7Dreport/chapter2.htm> (visited on 10/14/2020).

Filipiak, Wojciech et al. (01/2013). "Volatile Organic Compounds (VOCs) Released by Pathogenic Microorganisms in vitro: Potential Breath Biomarkers for Early-Stage Diagnosis of Disease". In: *Volatile Biomarkers.* Elsevier B.V., pp. 463–512. DOI: 10.1016/B978-0-44-462613-4.00023-4.

Glen, Stephanie (2017). *U Statistic: Definition, Different Types; Theil's U - Statistics How To.* <https://www.statisticshowto.com/u-statistic-theils/> (visited on 10/19/2020).

— (2018). *Uncertainty Coefficient: Definition.* <https://www.statisticshowto.com/uncertainty-coefficient/> (visited on 10/19/2020).

Granger, C. W.J. and P. Newbold (03/1973). "Some comments on the evaluation of economic forecasts". In: *Applied Economics* 5.1, pp. 35–47. DOI: 10.1080/00036847300000003.

Harris, Richard (2019). *'Statistical Significance' Is Overused And Often Misleading : Shots - Health News : NPR.* <https://www.npr.org/sections/health-shots/2019/03/20/705191851/statisticians-call-to-arms-reject-significance-and-embrace-uncertainty?t=1600534742009%7B%5C&%7Dt=1605086473289> (visited on 11/11/2020).

Hedayatpanah Shaldehi, Ahmad (2013). *Using Eta (η) correlation ratio in analyzing strongly nonlinear relationship between two Variables in Practical researches.* Tech. rep., pp. 213–220.

Hemmerich, W. (2019). *Cramer's V | StatistikGuru.* <https://statistikguru.de/lexikon/cramers-v.html> (visited on 10/26/2020).

HERE (2020). *HERE Technologies | The world's #1 location platform.* <https://www.here.com/> (visited on 08/22/2020).

Herz, R, H G SCHLICHTER, and W SIEGENER (1992). "ANGEWANDTE STATISTIK FUER VERKEHRS- UND REGIONALPLANER". In.

Hinden, Robert M. and Stephen E. Deering (2006). "IP Version 6 Addressing Architecture". In.

Hoang, Anh (2019). *Correlation between discrete (categorical) variables.* <https://rpubs.com/hoanganhngo610/558925> (visited on 10/19/2020).

Hucker, Marius (2020). *Tree algorithms explained: Ball Tree Algorithm vs. KD Tree vs. Brute Force.* <https://towardsdatascience.com/tree-algorithms-explained-ball-tree-algorithm-vs-kd-tree-vs-brute-force-9746debcd940> (visited on 10/30/2020).

Imielinski, T. and J. Navas (1996). "GPS-Based Addressing and Routing". In.

INDRIX (2015). *Floating-Car-Daten verbessern Genauigkeit der Verkehrsdaten in Bayern.* <https://inrix.com/press-releases/bavaria-quality-report-german/> (visited on 06/28/2020).

IZM (2020). *Wie entwickelt sich die Mobilfunknutzung weltweit?* <https://www.informationszentrum-mobilfunk.de/wirtschaft-gesellschaft/wirtschaftsfaktor/mobilfunknutzung-international> (visited on 10/13/2020).

J. Sheskin, David (1997). *Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition - Google Books.* <https://books.google.de/books?hl=de%7B%5C&%7Dlr=%7B%5C&%7Did=nvDqDwAAQBAJ%7B%5C&%7Doi=fnd%7B%5C&%7Dpg=PR19%7B%5C&%7Ddq=Sheskin,+David+J.+(1997).+Handbook+of+Parametric+and+Nonparametric+Statistical+Procedures.+Boca+Raton,+Fl:+CRC+Press.%7B%5C&%7Dots=JYvQRxpyfJ%7B%5C&%7Dsig=4yjXvDENCLzn4sQyj3w6mxm3K5E%7B%5C#%7Dv=onepage%7B%5C&%7Dq=Shes> (visited on 09/17/2020).

Jun, Ye (2020). *Everything You Need To Know About Correlation | Jun's blog.* <https://junye0798.com/post/everythin-you-need-to-know-about-correlation/> (visited on 10/19/2020).

Kessler, Lisa, Barbara Karl, and Klaus Bogenberger (2020). "Congestion Hot Spot Identification using Automated Pattern Recognition". In: 2020, pp. 1711–1717.

Laerd (2020). *Understanding the different types of variable in statistics | Laerd Statistics.* <https://statistics.laerd.com/statistical-guides/types-of-variable.php> (visited on 09/17/2020).

Lakens, Daniël (11/2013). "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs". In: *Frontiers in Psychology* 4.NOV, p. 863. DOI:

10.3389/fpsyg.2013.00863. <http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/abstract>.

LAPID (2018). *Häufige Stauursachen auf deutschen Straßen.* <https://blog.lapid.de/haeufige-stauursachen-auf-deutschen-stra%7B%5Css%7Den> (visited on 06/26/2020).

— (2020). *Staumeldesysteme und Floating Data (FCD/FPD).* <https://blog.lapid.de/staumeldesysteme-und-floating-data> (visited on 06/19/2020).

Leblanc, Vincent and Michael A A Cox (2017). "Interpretation of the point-biserial correlation coecient in the context of a school examination". In: 13.1. DOI: 10.20982/tqmp.13.1.p046.

Leon, Andrew C. (01/1998). "Descriptive and Inferential Statistics". In: *Comprehensive Clinical Psychology*. Elsevier, pp. 243–285. DOI: 10.1016/b0080-4270(73)00264-9.

McCue, Colleen (01/2007). "Data". In: *Data Mining and Predictive Analysis*. Elsevier, pp. 67–92. DOI: 10.1016/B978-075067796-7/50027-1. <https://linkinghub.elsevier.com/retrieve/pii/B9780750677967500271>.

Newson, Roger Benedict, Newson, and Roger (2002). "Parameters behind "nonparametric" statistics: Kendall's tau,Somers' D and median differences". In: *Stata Journal* 2.1, pp. 45–64. <https://econpapers.repec.org/RePEc:tsj:stataj:v:2:y:2002:i:1:p:45-64>.

ONS (2020). *Uncertainty and how we measure it for our surveys - Office for National Statistics.* <https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/uncertaintyandhowwemeasureit> (visited on 11/11/2020).

Outside Two Standard Deviations (2018). *An overview of correlation measures between categorical and continuous variables | by Outside Two Standard Deviations | Medium.* <https://medium.com/@outside2SDs/an-overview-of-correlation-measures-between-categorical-and-continuous-variables-4c7f85610365> (visited on 09/07/2020).

— (2020). *An overview of correlation measures between categorical and continuous variables.* <https://medium.com/@outside2SDs/an-overview-of-correlation-measures-between-categorical-and-continuous-variables-4c7f85610365> (visited on 10/17/2020).

Pearson, Karl (07/1900). "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50.302, pp. 157–175. DOI: 10.1080/14786440009463897. <http://www.economics.soton.ac.uk/staff/aldrich/1900.pdf>.

Piegorsch, Walter W (2002). *TABLES OF P-VALUES FOR t-AND CHI-SQUARE REFERENCE DISTRIBUTIONS*. Tech. rep.

Potvin, Pascal (2020). *CorrelationMatrixClustering/CorrelationMatrixClustering.ipynb at master · TheLoneNut/CorrelationMatrixClustering.* <https://github.com/TheLoneNut/CorrelationMatrixClustering/blob/master/CorrelationMatrixClustering.ipynb> (visited on 09/20/2020).

Prado, Kelvin Salton do (2017). *How DBSCAN works and why should we use it?* <https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80> (visited on 10/19/2020).

Ramzai, Juhi (2020). *Clearly explained: Pearson V/S Spearman Correlation Coefficient | Towards Data Science.* <https://towardsdatascience.com/clearly-explained-pearson-v-s-spearman-correlation-coefficient-ada2f473b8> (visited on 08/22/2020).

Randelhoff, Martin (2016). *Datenbasierte Verkehrsplanung: Vom Papier-Erhebungsbogen zu Big Data - Zukunft Mobilität.* <https://www.zukunft-mobilitaet.net/103615/analyse/datenbasierte-verkehrsplanung-big-data-mobilfunkdaten-optimierung-trajektorien/> (visited on 06/20/2020).

Regber, Linda (2016). *Effektstärke, Effektgröße & Effektstärkemaß berechnen | ANLEITUNG.* <http://lindaregber.com/effekt-berechnen-interpretieren/> (visited on 10/18/2020).

Reiter (2015). *Kendalls Tau.* <http://www.reiter1.com/Glossar/Kendall%7B%5C_%7DTau.htm> (visited on 10/18/2020).

Rempe, Felix (2018). "Traffic Speed Estimation and Prediction Using Floating Car Data". PhD thesis.

Retallack, Angus Eugene and Bertram Ostendorf (2019). *Current understanding of the effects of congestion on traffic accidents.* DOI: 10.3390/ijerph16183400. </pmc/articles/PMC6766193/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6766193/>.

SAGE (2014). *Learn to Use the Eta Coefficient Test in R With Data From the NIOSH Quality of Worklife Survey Learn to Use the Eta Coefficient Test in R With Data From the NIOSH Quality of Worklife Survey (2014) Student Guide.* Tech. rep.

Shinar, David and Richard Compton (05/2004). "Aggressive driving: An observational study of driver, vehicle, and situational variables". In: *Accident Analysis and Prevention* 36.3, pp. 429–437. DOI: 10.1016/S0001-4575(03)00037-X.

Sir Arthur Lewis Institute (2012). "Nominal-by-Interval Association Eta , the Correlation Ratio". In: *Quantitative Methods in Public Administration*, pp. 9–12.

SPSS (2020a). *Correlation (Pearson, Kendall, Spearman) - Statistics Solutions.* <https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/> (visited on 08/22/2020).

— (2020b). *Correlation in SPSS - Statistics Solutions.* <https://www.statisticssolutions.com/correlation-in-spss/> (visited on 08/22/2020).

StackExchange (2020). *Correlation - When to use Theil's U and Cramer's V, the danger of symmetrical data? - Cross Validated.* <https://stats.stackexchange.com/questions/443878/when-to-use-theils-u-and-cramers-v-the-danger-of-symmetrical-data> (visited on 09/07/2020).

Stmi (2020). *Bayerische Verkehrsunfallstatistik 2019 - Bayerisches Staatsministerium des Innern, für Sport und Integration.* <https://www.stmi.bayern.de/med/aktuell/archiv/2020/200220-verkehrsunfallstatistik-2019/index.php> (visited on 09/14/2020).

Sulthan, A (2018). *Is Pearson's Correlation coefficient appropriate for non-normal data?* <https://www.researchgate.net/post/Is%7B%5C_%7DPearsons%7B%5C_%7DCorrelation%7B%5C_%7Dcoefficient%7B%5C_%7Dappropriate%7B%5C_%7Dfor%7B%5C_%7Dnon-normal%7B%5C_%7Ddata> (visited on 09/28/2020).

Sun, Jian et al. (2016). "Analysis of safety factors for urban expressways considering the effect of congestion in Shanghai, China". In: *Accident Analysis and Prevention* 95, pp. 503–511. DOI: 10.1016/j.aap.2015.12.011.

Tang, Qing and Xianbiao Hu (2019). "Triggering behavior changes with information and incentives: An active traffic and demand management-oriented review". In: vol. 3. Academic Press, pp. 209–250. DOI: 10.1016/bs.atpp.2019.05.002.

Tate, Robert F (1954). *Correlation Between a Discrete and a Continuous Variable. Point-Biserial Correlation.* Tech. rep. 3, pp. 603–607.

Tenny, Steven and Ibrahim Abdelgawad (2020). *Statistical Significance.* <https://www.ncbi.nlm.nih.gov/books/NBK459346/> (visited on 11/11/2020).

Trafficon (2017). *ArbIS - Arbeitsstellenintegrationssystem - Betreuung und Datenbankpflege.* <https://www.trafficon.eu/projekte/arbis-arbeitsstellenintegrationssystem/> (visited on 06/24/2020).

TRB (2003). *Providing a Highway System with Reliable Travel Times | Blurbs New | Blurbs | Main.* <http://www.trb.org/Main/Blurbs/153181.aspx> (visited on 10/14/2020).

Walker, David A (2005). "JMASM19: A SPSS Matrix For Determining Effect Sizes From Three Categories: r And Functions Of r, Differences Between Proportions, And Standardized Differences Between Means". In: *Journal of Modern Applied Statistical Methods* 4.1, pp. 333–342. DOI: 10.22237/jmasm/1114907400. <http://digitalcommons.wayne.edu/jmasmhttp://digitalcommons.wayne.edu/jmasm/vol4/iss1/30>.

Wolfe, Douglas A. and Grant Schneider (2017). *Intuitive Introductory Statistics.* Springer Texts in Statistics. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-56072-4. <http://link.springer.com/10.1007/978-3-319-56072-4>.

Yildirim, Soner (2020). *DBSCAN Clustering — Explained. Detailed theorotical explanation and scikit-learn implementation.* <https://towardsdatascience.com/dbscan-clustering-explained-97556a2ad556> (visited on 10/11/2020).

Zhao, Yuzhen, Xiyu Liu, and Xiufeng Li (12/2018). "An improved DBSCAN algorithm based on cell-like P systems with promoters and inhibitors". In: *PLOS ONE* 13.12. Ed. by Manabu Sakakibara, e0200751. DOI: 10.1371/journal.pone.0200751. <https://dx.plos.org/10.1371/journal.pone.0200751>.

Zychlinski, Shaked (2018). *The Search for Categorical Correlation | Towards Data Science.* <https://towardsdatascience.com/the-search-for-categorical-correlation-a1cf7f1888c9> (visited on 09/19/2020).

Appendices

A BAYSIS Figures, Tables and Listings

	Strasse	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
Strasse	1.00	0.07	0.11	0.08	0.09	0.05	0.07	0.04	0.08	0.07	0.12	0.10	0.05	0.06	0.10	0.06	0.15	0.09	0.05	0.05
Kat	0.07	1.00	0.16	0.18	0.31	0.10	0.08	0.05	0.12	0.02	0.05	0.03	0.02	0.04	0.05	0.02	0.08	0.04	0.03	0.05
Typ	0.11	0.16	1.00	0.31	0.56	0.06	0.26	0.06	0.25	0.06	0.15	0.09	0.09	0.20	0.33	0.12	0.16	0.08	0.05	0.09
Betei	0.08	0.18	0.31	1.00	0.28	0.06	0.14	0.20	0.24	0.04	0.07	0.07	0.09	0.09	0.25	0.10	0.08	0.07	0.06	0.06
UArt1	0.09	0.31	0.56	0.28	1.00	0.08	0.21	0.05	0.32	0.05	0.14	0.09	0.10	0.22	0.25	0.09	0.16	0.08	0.05	0.06
UArt2	0.05	0.10	0.06	0.06	0.08	1.00	0.06	0.03	0.15	0.04	0.03	0.05	0.03	0.05	0.08	0.04	0.04	0.04	0.01	0.04
AUrs1	0.07	0.08	0.26	0.14	0.21	0.06	1.00	0.20	0.16	0.05	0.08	0.09	0.12	0.13	0.66	0.77	0.05	0.09	0.04	0.15
AUrs2	0.04	0.05	0.06	0.20	0.05	0.03	0.20	1.00	0.06	0.04	0.03	0.06	0.04	0.03	0.12	0.33	0.03	0.03	0.03	0.05
AufHi	0.08	0.12	0.25	0.24	0.32	0.15	0.16	0.06	1.00	0.04	0.08	0.10	0.08	0.10	0.25	0.09	0.06	0.07	0.04	0.06
Alkoh	0.07	0.02	0.06	0.04	0.05	0.04	0.05	0.04	0.04	1.00	0.02	0.00	0.11	0.11	0.03	0.01	0.05	0.08	0.01	0.05
Char1	0.12	0.05	0.15	0.07	0.14	0.03	0.08	0.03	0.08	0.02	1.00	0.58	0.04	0.05	0.10	0.03	0.06	0.03	0.02	0.04
Char2	0.10	0.03	0.09	0.07	0.09	0.05	0.09	0.06	0.10	0.00	0.58	1.00	0.04	0.04	0.08	0.03	0.08	0.03	0.02	0.04
Lich1	0.05	0.02	0.09	0.09	0.10	0.03	0.12	0.04	0.08	0.11	0.04	0.04	1.00	0.71	0.16	0.06	0.05	0.04	0.03	0.21
Lich2	0.06	0.04	0.20	0.09	0.22	0.05	0.13	0.03	0.10	0.11	0.05	0.04	0.71	1.00	0.16	0.06	0.17	0.04	0.03	0.20
Zust1	0.10	0.05	0.33	0.25	0.25	0.08	0.66	0.12	0.25	0.03	0.10	0.08	0.16	0.16	1.00	0.17	0.06	0.12	0.05	0.37
Zust2	0.06	0.02	0.12	0.10	0.09	0.04	0.77	0.33	0.09	0.01	0.03	0.03	0.06	0.06	0.17	1.00	0.05	0.06	0.02	0.17
Fstf	0.15	0.08	0.16	0.08	0.16	0.04	0.05	0.03	0.06	0.05	0.06	0.08	0.05	0.17	0.06	0.05	1.00	0.03	0.02	0.04
WoTag	0.09	0.04	0.08	0.07	0.08	0.04	0.09	0.03	0.07	0.08	0.03	0.03	0.04	0.04	0.12	0.06	0.03	1.00	0.13	0.09
FeiTag	0.05	0.03	0.05	0.06	0.05	0.01	0.04	0.03	0.04	0.01	0.02	0.02	0.03	0.03	0.05	0.02	0.02	0.13	1.00	0.13
Month	0.05	0.05	0.09	0.06	0.06	0.04	0.15	0.05	0.06	0.05	0.04	0.04	0.21	0.20	0.37	0.17	0.04	0.09	0.13	1.00

Table A.1: Correlation matrix for BAYSIS dataset, calculated with Cramer's V, η, τ, r_{pq}, r

	Strasse	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
Strasse	nan	0.000	0.000	0.000	0.000	0.000	0.000	0.871	0.000	0.000	0.000	0.000	0.165	0.000	0.000	0.006	0.000	0.000	0.073	0.002
Kat	0.000	nan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.225	0.000	0.058	0.260	0.000	0.000	0.068	0.000	0.000	0.069	0.000
Typ	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Betei	0.000	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.000	0.463	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UArt1	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
UArt2	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.045	0.000	0.017	0.043	0.001	0.122	0.001	0.000	0.002	0.000	0.016	0.998	0.054
AUrs1	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.157	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.386	0.000
AUrs2	0.871	0.000	0.000	0.000	0.000	0.045	0.000	nan	0.000	0.070	0.130	0.000	0.028	0.647	0.000	0.000	0.247	0.100	0.199	0.000
AufHi	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000
Alkoh	0.000	0.225	0.000	0.463	0.000	0.017	0.157	0.070	0.026	nan	0.689	0.754	0.000	0.000	0.035	0.745	0.004	0.000	0.279	0.017
Char1	0.000	0.000	0.000	0.000	0.000	0.043	0.000	0.130	0.000	0.689	nan	0.000	0.000	0.000	0.000	0.017	0.000	0.007	0.415	0.084
Char2	0.000	0.058	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.754	0.000	nan	0.001	0.000	0.000	0.012	0.000	0.250	0.020	0.075
Lich1	0.165	0.260	0.000	0.000	0.000	0.122	0.000	0.028	0.000	0.000	0.000	0.001	nan	0.000	0.000	0.000	0.000	0.002	0.015	0.000
Lich2	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.647	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.000	0.001	0.033	0.000
Zust1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.035	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.000
Zust2	0.006	0.068	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.745	0.017	0.012	0.000	0.000	0.000	nan	0.000	0.000	0.125	0.000
Fstf	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.247	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.033	0.660	0.081
WoTag	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.100	0.000	0.000	0.007	0.250	0.002	0.001	0.000	0.000	0.033	nan	0.000	0.000
FeiTag	0.073	0.069	0.000	0.000	0.001	0.998	0.386	0.199	0.066	0.279	0.415	0.020	0.015	0.033	0.000	0.125	0.660	0.000	nan	0.000
Month	0.002	0.000	0.000	0.000	0.000	0.054	0.000	0.000	0.000	0.017	0.084	0.075	0.000	0.000	0.000	0.000	0.081	0.000	0.000	nan

Table A.2: Significancy matrix for BAYSIS dataset

	Strasse	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
Strasse	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Kat	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Typ	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Betei	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt1	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt2	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs1	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs2	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V
AufHi	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V
Alkoh	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V
Char1	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V
Char2	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V
Lich1	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V
Lich2	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V
Zust1	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V
Zust2	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V
Fstf	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V
WoTag	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V
FeiTag	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V
Month	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN

Table A.3: Coefficient matrix for BAYSIS dataset

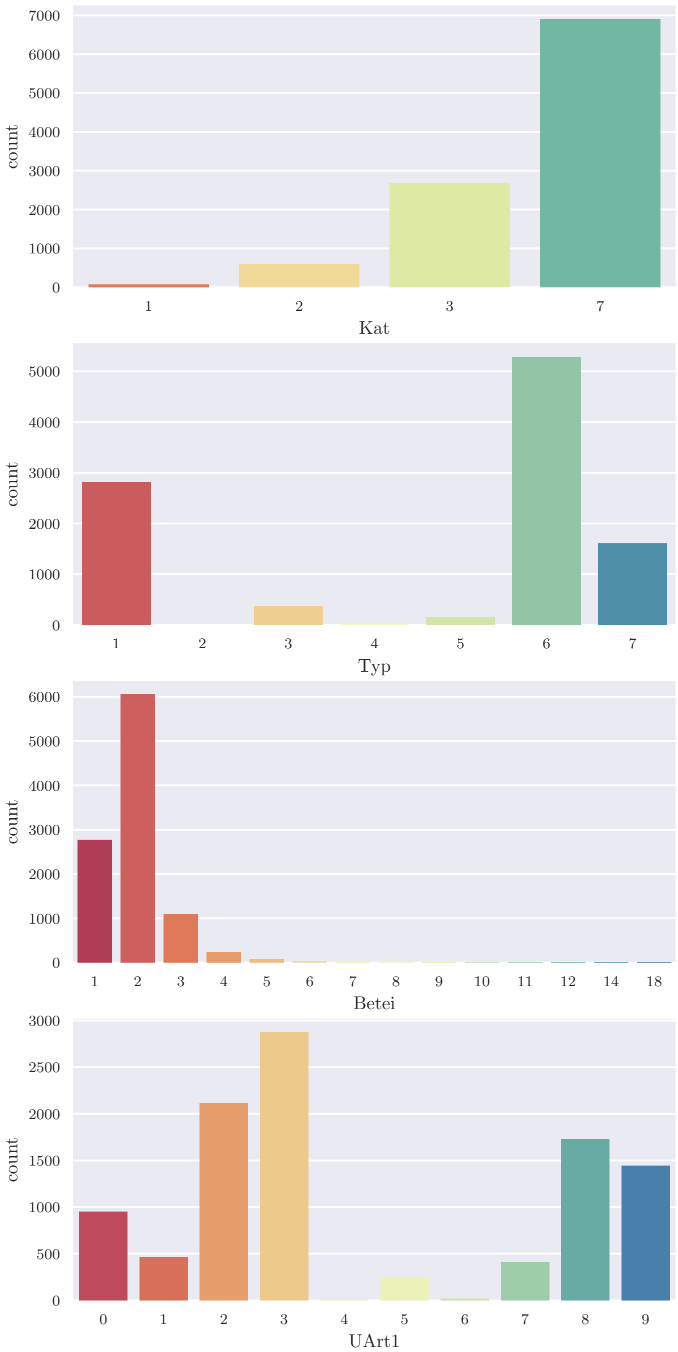


Figure A.1: Distribution of the accident category Kat, Typ, Betei and UArt

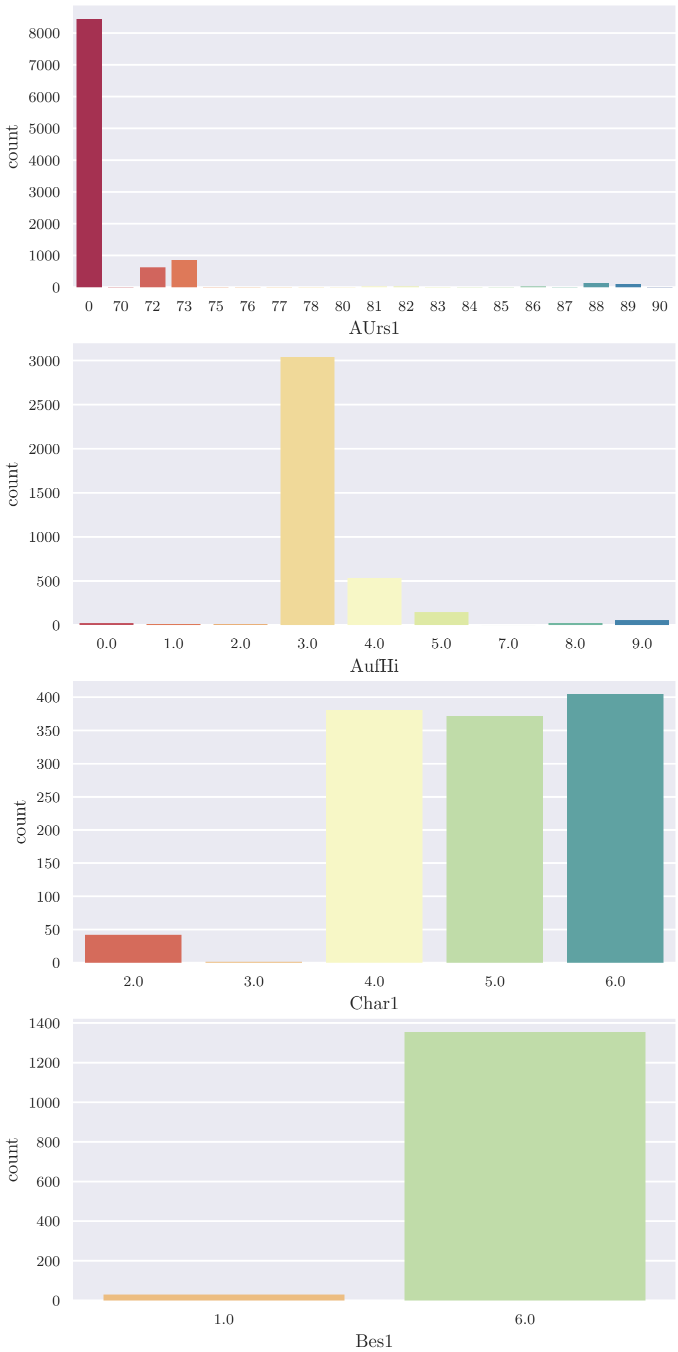


Figure A.2: Distribution of the accident category AUrs, AufHi, Char and Bes

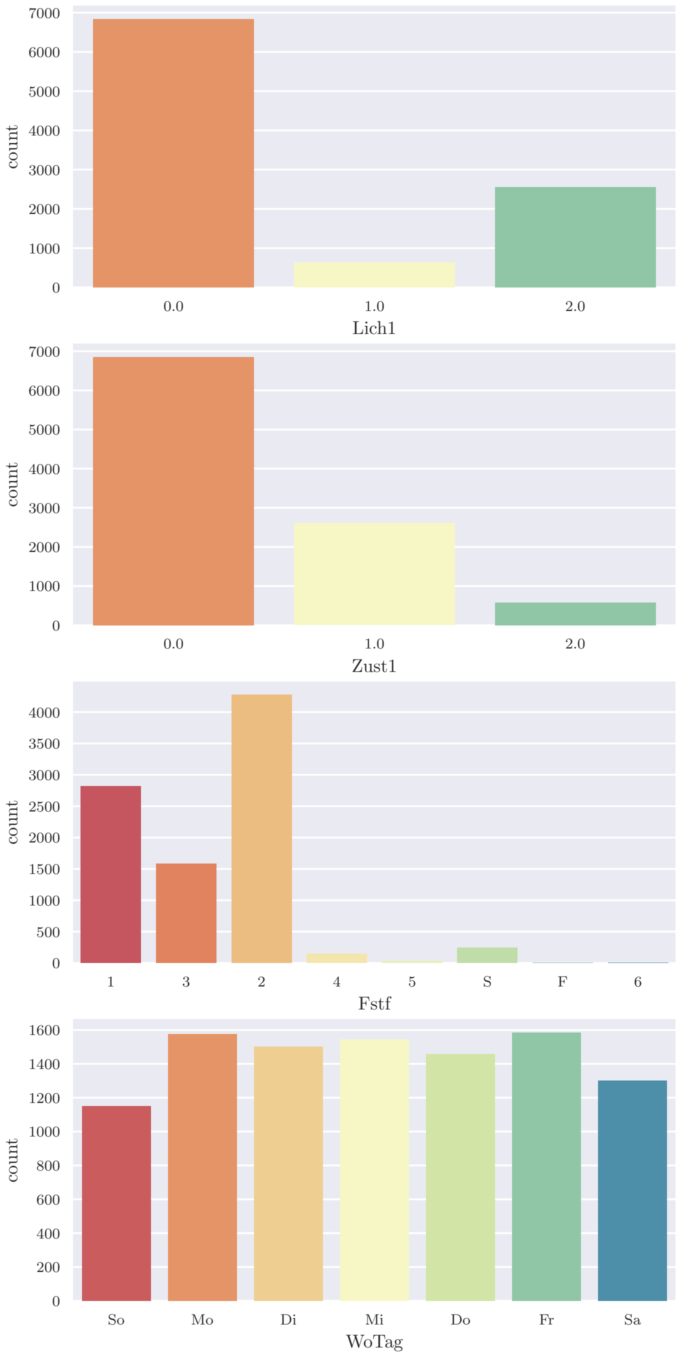


Figure A.3: Distribution of the accident category Lich, Zust, Fstf and WoTag

	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A6	0.00															
A9	0.01	1.00														
A70	0.03	1.00	1.00													
A96	0.00	1.00	0.27	1.00												
A7	0.00	1.00	1.00	1.00	1.00											
A73	0.00	1.00	0.31	1.00	1.00	1.00										
A99	1.00	1.00	1.00	1.00	0.50	1.00	0.59									
A92	0.00	1.00	0.16	1.00	1.00	1.00	1.00	0.22								
A93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A94	0.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.4: Pairwise Wilcoxon T-test for *Street* and *Maximal Temporal Extent* complete

	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A6	0.84															
A9	0.36	1.00														
A70	1.00	1.00	1.00													
A96	0.10	1.00	1.00	1.00												
A7	1.00	1.00	1.00	1.00	1.00											
A73	0.00	1.00	0.51	1.00	1.00	1.00										
A99	0.02	1.00	1.00	1.00	1.00	1.00	1.00									
A92	0.26	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
A93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A94	0.28	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.5: Pairwise Wilcoxon T-test for *Street* and *Average Temporal Extent* complete

	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A6	0.40															
A9	0.00	1.00														
A70	0.00	0.83	0.54													
A96	0.00	1.00	0.14	1.00												
A7	0.00	1.00	1.00	1.00	1.00											
A73	0.00	0.00	0.00	1.00	1.00	0.59										
A99	1.00	1.00	1.00	0.80	0.31	1.00	0.00									
A92	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00								
A93	0.03	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A94	0.00	0.11	0.03	1.00	1.00	1.00	1.00	0.09	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.6: Pairwise Wilcoxon T-test for *Street* and *Maximal Spatial Extent* complete

	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A6	1.00															
A9	1.00	1.00														
A70	0.05	0.83	0.71													
A96	0.05	1.00	1.00	1.00												
A7	1.00	1.00	1.00	1.00	1.00											
A73	0.00	0.00	0.00	1.00	0.00	0.00										
A99	0.00	1.00	0.06	1.00	1.00	1.00	0.51									
A92	0.00	0.61	0.03	1.00	1.00	1.00	1.00	1.00								
A93	0.03	0.46	0.16	1.00	1.00	1.00	1.00	1.00	1.00							
A94	0.00	0.07	0.01	1.00	0.36	0.31	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.7: Pairwise Wilcoxon T-test for *Street* and *Average Spatial Extent* complete

	A3	A6	A9	A70	A96	A7	A73	A99	A92	A93	A94	A72	A995	A95	A71	A45
A6	0.05															
A9	0.00	1.00														
A70	1.00	1.00	1.00													
A96	0.00	1.00	0.00	1.00												
A7	0.00	1.00	0.01	1.00	1.00											
A73	0.04	1.00	1.00	1.00	1.00	1.00										
A99	0.88	0.00	0.00	0.09	0.00	0.00	0.00									
A92	0.00	1.00	0.12	1.00	1.00	1.00	1.00	0.00								
A93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.13	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.8: Pairwise Wilcoxon T-test for *Street* and *Coverage* complete

	0	1	2	3	4	5	6	7	8
1	1.00								
2	1.00	1.00							
3	1.00	1.00	1.00						
4	0.57	0.22	0.31	0.17					
5	0.04	0.17	0.00	0.00	0.01				
6	0.40	0.19	0.23	0.16	1.00	0.01			
7	0.00	0.00	0.00	0.00	1.00	0.00	1.00		
8	0.02	0.00	0.00	0.00	1.00	0.00	1.00	0.32	
9	0.01	0.00	0.00	0.00	1.00	0.00	1.00	0.50	1.00

Table A.9: Pairwise Wilcoxon T-test for *UArt* and *Temporal Distance* complete

	0	1	2	3	4	5	6	7	8
1	1.00								
2	1.00	1.00							
3	1.00	1.00	1.00						
4	1.00	1.00	1.00	1.00					
5	0.41	0.10	0.00	0.01	0.65				
6	1.00	1.00	1.00	1.00	1.00	1.00			
7	0.30	0.36	0.12	0.05	1.00	0.00	1.00		
8	0.01	0.01	0.00	0.00	1.00	0.00	1.00	1.00	
9	0.05	0.07	0.00	0.00	1.00	0.00	1.00	1.00	1.00

Table A.10: Pairwise Wilcoxon T-test for *UArt* and *Coverage* complete

	0	72	73	75	76	77	80	81	82	83	84	86	87	88
72	1.00													
73	0.14	0.16												
75	1.00	1.00	1.00											
76	1.00	1.00	1.00	1.00										
77	1.00	1.00	1.00	1.00	1.00									
80	1.00	1.00	1.00	1.00	1.00	1.00								
81	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33						
83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
88	0.03	0.03	0.18	1.00	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00	
89	1.00	1.00	0.06	1.00	1.00	1.00	1.00	0.71	1.00	1.00	1.00	1.00	1.00	0.10

Table A.11: Pairwise Wilcoxon T-test for *AUrs1* and *Average Spatial Extent*

	0	72	73	75	76	77	80	81	82	83	84	86	87	88
72	1.00													
73	0.00	1.00												
75	1.00	1.00	1.00											
76	1.00	1.00	1.00	1.00										
77	1.00	1.00	1.00	1.00	1.00									
80	1.00	1.00	1.00	1.00		1.00								
81	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
86	0.57	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
88	0.08	1.00	0.54	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.12: Pairwise Wilcoxon T-test for *AUrs1* and *Temporal Distance*

	0	72	73	75	76	77	80	81	82	83	84	86	87	88
72	0.02													
73	0.00	1.00												
75	1.00	1.00	1.00											
76	1.00	1.00	1.00	1.00										
77	1.00	1.00	1.00	1.00	1.00									
80	1.00	1.00	1.00	1.00	1.00	1.00								
81	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
86	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
88	0.46	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	1.00	1.00

Table A.13: Pairwise Wilcoxon T-test for *AUrs* and *Coverage*

	0	1	2	3	4	5	8
1	1.00						
2	1.00	1.00					
3	0.00	1.00	1.00				
4	0.03	1.00	1.00	1.00			
5	0.28	1.00	1.00	1.00	1.00		
8	1.00	1.00	1.00	1.00	1.00	1.00	
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.14: Pairwise Wilcoxon T-test for *AufHi* and *Temporal Distance* complete

	0	1	2	3	4	5	8
1	1.00						
2	1.00	1.00					
3	0.00	1.00	1.00				
4	0.51	1.00	1.00	1.00			
5	1.00	1.00	1.00	1.00	1.00		
8	1.00	1.00	1.00	1.00	1.00	1.00	
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.15: Pairwise Wilcoxon T-test for *AufHi* and *Coverage* complete

	Di	Mi	Do	Fr	Sa	So	Mo
Mi	1.00						
Do	1.00	1.00					
Fr	1.00	1.00	1.00				
Sa	0.20	0.48	0.20	1.00			
So	0.00	0.02	0.00	0.14	1.00		
Mo	0.00	0.01	0.00	0.10	1.00	1.00	

Table A.16: Pairwise Wilcoxon T-test for *WoTag* and *Coverage* complete

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	1.00	0.82	0.58	0.49	-0.29	-0.09	-0.25	0.03	-0.01	0.25	0.15	0.06	0.09	0.12	0.08	0.14	0.06	0.14	-0.01	0.05	0.04	0.03	0.03	0.11	0.01	-0.00	0.10	-0.00	0.14
TAvg	0.82	1.00	0.30	0.51	-0.16	-0.07	0.10	0.02	-0.01	0.17	0.16	0.06	0.08	0.09	0.05	0.11	0.08	0.14	0.02	0.01	0.02	0.03	0.02	0.06	0.05	0.01	0.09	-0.01	0.10
SMax	0.58	0.30	1.00	0.64	-0.27	-0.09	-0.50	0.01	-0.05	0.31	0.10	0.09	0.08	0.18	0.08	0.14	0.06	0.13	-0.04	0.05	0.03	0.08	0.07	0.08	0.05	0.03	0.08	-0.00	0.15
SAvg	0.49	0.51	0.64	1.00	-0.09	-0.13	0.10	0.01	-0.06	0.25	0.23	0.10	0.09	0.19	0.06	0.17	0.11	0.06	-0.02	0.04	0.00	0.02	0.02	0.01	0.12	0.04	0.11	0.01	0.11
TDist	-0.29	-0.16	-0.27	-0.09	1.00	0.05	0.31	-0.02	0.02	0.17	0.19	0.24	-0.04	0.29	0.11	0.20	0.11	0.21	0.04	0.11	0.11	0.13	0.14	0.13	0.02	0.03	0.10	0.01	0.10
SDist	-0.09	-0.07	-0.09	-0.13	0.05	1.00	-0.04	-0.02	0.02	0.22	0.10	0.05	-0.03	0.07	0.06	0.15	0.02	0.02	-0.03	0.07	0.02	0.05	0.05	0.05	0.01	0.04	0.08	-0.00	0.07
Cov	-0.25	0.10	-0.50	0.10	0.31	-0.04	1.00	-0.00	0.01	0.29	0.13	0.21	-0.02	0.24	0.12	0.22	0.10	0.21	0.07	0.11	0.08	0.15	0.15	0.17	0.03	0.02	0.16	0.01	0.14
TLCar	0.03	0.02	0.01	0.01	-0.02	-0.02	-0.00	1.00	0.00	0.08	0.02	0.06	0.01	0.08	0.05	0.08	0.05	0.04	-0.03	0.04	0.04	0.03	0.02	0.03	0.02	-0.03	0.04	-0.01	0.04
TLHGV	-0.01	-0.01	-0.05	-0.06	0.02	0.02	0.01	0.00	1.00	0.11	0.02	0.06	-0.02	0.06	0.04	0.14	0.08	0.06	-0.00	0.05	0.02	0.03	0.03	0.06	0.02	-0.01	0.07	0.04	0.12
Str	0.25	0.17	0.31	0.25	0.17	0.22	0.29	0.08	0.11	1.00	0.13	0.13	0.10	0.11	0.11	0.10	0.06	0.11	0.07	0.14	0.10	0.11	0.10	0.15	0.12	0.16	0.12	0.11	0.11
Kat	0.15	0.16	0.10	0.23	0.19	0.10	0.13	0.02	0.02	0.13	1.00	0.20	0.21	0.35	0.13	0.11	0.05	0.15	0.05	0.08	0.07	0.06	0.06	0.06	0.07	0.09	0.09	0.07	0.10
Typ	0.06	0.06	0.09	0.10	0.24	0.05	0.21	0.06	0.06	0.13	0.20	1.00	0.31	0.59	0.09	0.25	0.08	0.24	0.10	0.13	0.19	0.07	0.09	0.23	0.13	0.13	0.10	0.10	0.09
Betei	0.09	0.08	0.08	0.09	-0.04	-0.03	-0.02	0.01	-0.02	0.10	0.21	0.31	1.00	0.29	0.08	0.18	0.30	0.21	0.04	0.08	0.15	0.08	0.06	0.17	0.28	0.09	0.09	0.20	0.08
UArt1	0.12	0.09	0.18	0.19	0.29	0.07	0.24	0.08	0.06	0.11	0.35	0.59	0.29	1.00	0.13	0.22	0.10	0.31	0.10	0.16	0.17	0.08	0.08	0.20	0.09	0.14	0.11	0.08	0.08
UArt2	0.08	0.05	0.08	0.06	0.11	0.06	0.12	0.05	0.04	0.11	0.13	0.09	0.08	0.13	1.00	0.15	0.05	0.26	0.02	0.08	0.11	0.07	0.07	0.07	0.05	0.06	0.07	0.04	0.08
AUrs1	0.14	0.11	0.14	0.17	0.20	0.15	0.22	0.08	0.14	0.10	0.11	0.25	0.18	0.22	0.15	1.00	0.45	0.16	0.03	0.10	0.15	0.13	0.10	0.59	0.49	0.07	0.10	0.35	0.14
AUrs2	0.06	0.08	0.06	0.11	0.11	0.02	0.10	0.05	0.08	0.06	0.05	0.08	0.30	0.10	0.05	0.45	1.00	0.04	0.01	0.04	0.11	0.06	0.03	0.20	0.45	0.03	0.07	0.32	0.08
AufHi	0.14	0.14	0.13	0.06	0.21	0.02	0.21	0.04	0.06	0.11	0.15	0.24	0.21	0.31	0.26	0.16	0.04	1.00	0.02	0.09	0.18	0.06	0.06	0.19	0.06	0.09	0.09	0.07	0.08
Alkoh	-0.01	0.02	-0.04	-0.02	0.04	-0.03	0.07	-0.03	-0.00	0.07	0.05	0.10	0.04	0.10	0.02	0.03	0.01	0.02	1.00	0.06	0.00	0.13	0.11	0.02	0.01	0.06	0.05	0.02	0.06
Char1	0.05	0.01	0.05	0.04	0.11	0.07	0.11	0.04	0.05	0.14	0.08	0.13	0.08	0.16	0.08	0.10	0.04	0.09	0.06	1.00	0.59	0.07	0.06	0.09	0.03	0.06	0.06	0.03	0.08
Char2	0.04	0.02	0.03	0.00	0.11	0.02	0.08	0.04	0.02	0.10	0.07	0.19	0.15	0.17	0.11	0.15	0.11	0.18	0.00	0.59	1.00	0.07	0.07	0.12	0.00	0.12	0.09	0.05	0.07
Lich1	0.03	0.03	0.08	0.02	0.13	0.05	0.15	0.03	0.03	0.11	0.06	0.07	0.08	0.08	0.07	0.13	0.06	0.06	0.13	0.07	0.07	1.00	0.71	0.15	0.03	0.06	0.07	0.03	0.25
Lich2	0.03	0.02	0.07	0.02	0.14	0.05	0.15	0.02	0.03	0.10	0.06	0.09	0.06	0.08	0.07	0.10	0.03	0.06	0.11	0.06	0.07	0.71	1.00	0.15	0.01	0.05	0.07	0.02	0.25
Zust1	0.11	0.06	0.08	0.01	0.13	0.05	0.17	0.03	0.06	0.15	0.06	0.23	0.17	0.20	0.07	0.59	0.20	0.19	0.02	0.09	0.12	0.15	0.15	1.00	0.18	0.07	0.09	0.13	0.33
Zust2	0.01	0.05	0.05	0.12	0.02	0.01	0.03	0.02	0.02	0.12	0.07	0.13	0.28	0.09	0.05	0.49	0.45	0.06	0.01	0.03	0.00	0.03	0.01	0.18	1.00	0.02	0.07	1.00	0.22
Fstf	-0.00	0.01	0.03	0.04	0.03	0.04	0.02	-0.03	-0.01	0.16	0.09	0.13	0.09	0.14	0.06	0.07	0.03	0.09	0.06	0.06	0.12	0.06	0.05	0.07	0.02	1.00	0.06	0.04	0.08
WoTag	0.10	0.09	0.08	0.11	0.10	0.08	0.16	0.04	0.07	0.12	0.09	0.10	0.09	0.11	0.07	0.10	0.07	0.09	0.05	0.06	0.09	0.07	0.07	0.09	0.07	0.06	1.00	0.12	0.12
FeiTag	-0.00	-0.01	-0.00	0.01	0.01	-0.00	0.01	-0.01	0.04	0.11	0.07	0.10	0.20	0.08	0.04	0.35	0.32	0.07	0.02	0.03	0.05	0.03	0.02	0.13	1.00	0.04	0.12	1.00	0.20
Month	0.14	0.10	0.15	0.11	0.10	0.07	0.14	0.04	0.12	0.11	0.10	0.09	0.08	0.08	0.08	0.14	0.08	0.08	0.06	0.08	0.07	0.25	0.25	0.33	0.22	0.08	0.12	0.20	1.00

Table A.17: Correlation matrix for BAYSIS matched data, calculated with Cramer's V, η, τ, r_{pq}, r

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei
TMax	nan	0.000	0.000	0.000	0.000	0.000	0.000	0.211	0.548	0.000	0.000	0.000	0.000
TAvg	0.000	nan	0.000	0.000	0.000	0.003	0.000	0.431	0.792	0.000	0.000	0.000	0.000
SMax	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.681	0.034	0.000	0.000	0.000	0.000
SAvg	0.000	0.000	0.000	nan	0.000	0.000	0.000	0.613	0.009	0.000	0.000	0.000	0.000
TDist	0.000	0.000	0.000	0.000	nan	0.022	0.000	0.361	0.381	0.000	0.000	0.000	0.027
SDist	0.000	0.003	0.000	0.000	0.022	nan	0.123	0.302	0.487	0.000	0.000	0.000	0.115
Cov	0.000	0.000	0.000	0.000	0.000	0.123	nan	0.937	0.708	0.000	0.000	0.000	0.180
TLCar	0.211	0.431	0.681	0.613	0.361	0.302	0.937	nan	0.941	0.000	0.000	0.000	0.438
TLHGV	0.548	0.792	0.034	0.009	0.381	0.487	0.708	0.941	nan	0.000	0.000	0.000	0.277
Str	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.304
Kat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000
Typ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000
Betei	0.000	0.000	0.000	0.000	0.027	0.115	0.180	0.438	0.277	0.304	0.000	0.000	nan
UArt1	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
UArt2	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.017	0.000	0.001	0.032
AUrs1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.009	0.000	0.000
AUrs2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.593	0.001	0.000
AufHi	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000
Alkoh	0.817	0.342	0.054	0.316	0.109	0.242	0.005	0.264	0.960	0.920	0.166	0.002	0.964
Char1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022
Char2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.339	0.036	0.000	0.000
Lich1	0.000	0.000	0.000	0.000	0.000	0.359	0.000	0.000	0.000	0.044	0.032	0.043	0.171
Lich2	0.000	0.000	0.000	0.000	0.000	0.337	0.000	0.000	0.000	0.229	0.049	0.003	0.558
Zust1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000
Zust2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.037	0.019	0.000	0.000
Fstf	0.894	0.672	0.094	0.012	0.169	0.069	0.221	0.083	0.705	0.000	0.001	0.000	0.000
WoTag	0.000	0.000	0.000	0.000	0.072	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000
FeiTag	0.802	0.720	0.825	0.618	0.730	0.982	0.732	0.736	0.063	0.115	0.009	0.000	0.000
Month	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.015	0.017	0.088

	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	0.000	0.000	0.000	0.000	0.000	0.817	0.000	0.000	0.000	0.000	0.000	0.000	0.894	0.000	0.802	0.000
TAvg	0.000	0.000	0.000	0.000	0.000	0.342	0.000	0.000	0.000	0.000	0.000	0.000	0.672	0.000	0.720	0.000
SMax	0.000	0.000	0.000	0.000	0.000	0.054	0.000	0.000	0.000	0.000	0.000	0.000	0.094	0.000	0.825	0.000
SAvg	0.000	0.000	0.000	0.000	0.000	0.316	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.618	0.000
TDist	0.004	0.000	0.000	0.000	0.000	0.109	0.000	0.000	0.000	0.000	0.000	0.000	0.169	0.072	0.730	0.000
SDist	0.000	0.006	0.000	0.000	0.000	0.242	0.000	0.000	0.359	0.337	0.000	0.000	0.069	0.000	0.982	0.000
Cov	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.221	0.000	0.732	0.000
TLCar	0.000	0.000	0.000	0.000	0.000	0.264	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	0.736	0.000
TLHGV	0.000	0.000	0.000	0.000	0.000	0.960	0.000	0.000	0.000	0.000	0.000	0.000	0.705	0.000	0.063	0.000
Str	0.003	0.017	0.004	1.000	0.017	0.920	0.000	0.339	0.044	0.229	0.000	0.037	0.000	0.000	0.115	0.001
Kat	0.000	0.000	0.009	0.593	0.000	0.166	0.000	0.036	0.032	0.049	0.030	0.019	0.001	0.003	0.009	0.015
Typ	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.043	0.003	0.000	0.000	0.000	0.000	0.000	0.017
Betei	0.000	0.032	0.000	0.000	0.000	0.964	0.022	0.000	0.171	0.558	0.000	0.000	0.000	0.000	0.000	0.088
UArt1	nan	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.092	0.110	0.000	0.136	0.000	0.000	0.128	0.078
UArt2	0.000	nan	0.000	0.997	0.000	0.999	0.033	0.003	0.296	0.344	0.227	0.871	0.529	0.309	0.977	0.310
AUrs1	0.000	0.000	nan	0.000	0.000	1.000	0.058	0.000	0.001	0.064	0.000	0.000	0.999	0.004	0.000	0.000
AUrs2	0.000	0.997	0.000	nan	1.000	1.000	0.926	0.002	0.446	0.977	0.000	0.000	1.000	0.069	0.000	0.281
AufHi	0.000	0.000	0.000	1.000	nan	0.994	0.000	0.000	0.372	0.539	0.000	0.361	0.000	0.000	0.329	0.256
Alkoh	0.018	0.999	1.000	1.000	0.994	nan	0.147	0.927	0.000	0.000	0.632	0.568	0.518	0.805	0.635	0.758
Char1	0.000	0.033	0.058	0.926	0.000	0.147	nan	0.000	0.038	0.090	0.001	0.865	0.345	0.357	0.949	0.212
Char2	0.000	0.003	0.000	0.002	0.000	0.927	0.000	nan	0.018	0.013	0.000	0.852	0.001	0.025	0.133	0.666
Lich1	0.092	0.296	0.001	0.446	0.372	0.000	0.038	0.018	nan	0.000	0.000	0.472	0.583	0.131	0.609	0.000
Lich2	0.110	0.344	0.064	0.977	0.539	0.000	0.090	0.013	0.000	nan	0.000	0.812	0.785	0.119	0.708	0.000
Zust1	0.000	0.227	0.000	0.000	0.000	0.632	0.001	0.000	0.000	0.000	nan	0.000	0.294	0.014	0.000	0.000
Zust2	0.136	0.871	0.000	0.000	0.361	0.568	0.865	0.852	0.472	0.812	0.000	nan	0.998	0.347	0.000	0.000
Fstf	0.000	0.529	0.999	1.000	0.000	0.518	0.345	0.001	0.583	0.785	0.294	0.998	nan	0.343	0.987	0.326
WoTag	0.000	0.309	0.004	0.069	0.000	0.805	0.357	0.025	0.131	0.119	0.014	0.347	0.343	nan	0.000	0.000
FeiTag	0.128	0.977	0.000	0.000	0.329	0.635	0.949	0.133	0.609	0.708	0.000	0.000	0.987	0.000	nan	0.000
Month	0.078	0.310	0.000	0.281	0.256	0.758	0.212	0.666	0.000	0.000	0.000	0.000	0.326	0.000	0.000	nan

Table A.18: Significancy matrix for BAYSIS matched data

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	NaN	r	r	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TAvg	r	NaN	r	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SMax	r	r	NaN	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SAvg	r	r	r	NaN	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TDist	r	r	r	r	NaN	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SDist	r	r	r	r	r	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Cov	r	r	r	r	r	r	NaN	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLCar	r	r	r	r	r	r	r	NaN	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLHGV	r	r	r	r	r	r	r	r	NaN	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Str	η	η	η	η	η	η	η	η	η	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Kat	η	η	η	η	η	η	η	η	η	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Typ	η	η	η	η	η	η	η	η	η	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Betei	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt1	η	η	η	η	η	η	η	η	η	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V
AufHi	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V
Alkoh	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V
Char1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V
Char2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V
Lich1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V
Lich2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V
Zust1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V
Zust2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V
Fstf	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V
WoTag	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V
FeiTag	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V
Month	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN

Table A.19: Coefficient matrix for BAYSIS matched data

	1	3	4	5	6
3	0.05				
4	1.00	0.38			
5	0.96	1.00	0.96		
6	0.00	0.96	0.51	1.00	
7	1.00	0.04	1.00	0.96	0.01

Table A.20: Pairwise Wilcoxon T-test for *Typ* and *Temporal Distance* (Jam Initiator) complete

	0	1	2	3	4	5	6	7	8
1	1.00								
2	1.00	1.00							
3	1.00	0.04	0.00						
4	1.00	1.00	1.00	1.00					
5	1.00	1.00	1.00	1.00	1.00				
6	1.00	1.00	1.00	1.00	1.00	1.00			
7	0.75	0.05	0.13	1.00	1.00	1.00	1.00		
8	1.00	0.21	0.47	1.00	1.00	1.00	1.00	1.00	
9	1.00	0.45	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.21: Pairwise Wilcoxon T-test for *UArt1* and *Maximal Temporal Extent* (Jam Initiator) complete

	0	1	2	3	4	5	6	7	8
1	1.00								
2	1.00	1.00							
3	1.00	1.00	1.00						
4	1.00	1.00	1.00	1.00					
5	1.00	1.00	1.00	1.00	1.00				
6	1.00	1.00	1.00	1.00	1.00	1.00			
7	1.00	0.06	0.00	0.00	1.00	0.10	1.00		
8	1.00	1.00	0.01	0.01	1.00	1.00	1.00	1.00	
9	1.00	1.00	0.05	0.01	1.00	0.78	1.00	0.87	1.00

Table A.22: Pairwise Wilcoxon T-test for *UArt1* and *Temporal Distance* (Jam Initiator) complete

	0	1	2	3	4	5	6	7	8
1	1.00								
2	1.00	1.00							
3	1.00	1.00	1.00						
4	1.00	1.00	1.00	1.00					
5	1.00	1.00	1.00	1.00	1.00				
6	1.00	1.00	1.00	1.00	1.00	1.00			
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
8	1.00	1.00	0.13	0.09	1.00	0.72	1.00	1.00	
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.23: Pairwise Wilcoxon T-test for *UArt1* and *Coverage* (Jam Initiator)

	72	73	75	77	81	82	83	84	86	87	88
72											
73	0.63										
75	1.00	1.00									
77	1.00	1.00	1.00								
81	1.00	1.00	1.00	1.00							
82	1.00	1.00	1.00	1.00	1.00						
83	1.00	1.00	1.00	1.00	1.00	1.00					
84	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
89	0.33	0.96	1.00	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00

Table A.24: Pairwise Wilcoxon T-test for *AUrs1* and *Coverage* (Jam Initiator)

	-1	0	1	2	3	4	5	8
0	1.00							
1	1.00	1.00						
2	1.00	1.00	1.00					
3	1.00	1.00	1.00	1.00				
4	1.00	1.00	1.00	1.00	1.00			
5	1.00	1.00	1.00	1.00	1.00	1.00		
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.25: Pairwise Wilcoxon T-test for *AufHi* and *Maximal Temporal Extent* (Jam Initiator)

	-1	0	1	2	3	4	5	8
0	1.00							
1	1.00	1.00						
2	1.00	1.00	1.00					
3	0.20	1.00	1.00	1.00				
4	0.25	1.00	1.00	1.00	1.00			
5	1.00	1.00	1.00	1.00	1.00	1.00		
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.26: Pairwise Wilcoxon T-test for *AufHi* and *Temporal Distance* (Jam Initiator)

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	1.00	0.89	0.60	0.53	-0.11	-0.03	-0.21	0.02	-0.00	0.18	0.29	0.06	0.14	0.16	0.05	0.23	0.15	0.16	0.00	0.05	0.04	0.02	0.02	0.10	0.10	0.03	0.07	-0.02	0.12
TAvg	0.89	1.00	0.39	0.55	-0.09	-0.03	0.08	0.01	-0.01	0.15	0.27	0.11	0.09	0.20	0.07	0.21	0.20	0.14	0.02	0.04	0.04	0.04	0.01	0.08	0.16	0.03	0.10	-0.01	0.11
SMax	0.60	0.39	1.00	0.72	-0.09	-0.01	-0.46	0.02	-0.00	0.26	0.20	0.08	0.11	0.20	0.07	0.22	0.11	0.13	-0.06	0.09	0.06	0.13	0.08	0.10	0.07	0.03	0.08	-0.02	0.14
SAvg	0.53	0.55	0.72	1.00	-0.01	-0.04	0.10	0.05	-0.04	0.24	0.26	0.16	0.08	0.27	0.06	0.25	0.19	0.09	-0.03	0.07	0.02	0.03	0.02	0.09	0.16	0.04	0.12	0.00	0.13
TDist	-0.11	-0.09	-0.09	-0.01	1.00	0.03	0.13	-0.01	0.06	0.10	0.10	0.21	-0.13	0.24	0.03	0.22	0.15	0.14	0.07	0.14	0.12	0.18	0.17	0.13	0.01	0.05	0.10	0.00	0.11
SDist	-0.03	-0.03	-0.01	-0.04	0.03	1.00	-0.07	-0.00	-0.00	0.06	0.07	0.03	-0.02	0.07	0.02	0.02	0.00	0.03	-0.00	0.01	0.01	0.02	0.02	0.02	0.00	0.07	0.09	-0.01	0.11
Cov	-0.21	0.08	-0.46	0.10	0.13	-0.07	1.00	0.02	-0.04	0.22	0.04	0.18	-0.07	0.19	0.06	0.25	0.13	0.14	0.08	0.11	0.07	0.12	0.09	0.18	0.04	0.03	0.13	0.03	0.16
TLCar	0.02	0.01	0.02	0.05	-0.01	-0.00	0.02	1.00	0.02	0.18	0.03	0.10	0.02	0.15	0.10	0.12	0.09	0.07	0.01	0.08	0.04	0.04	0.04	0.06	0.01	0.03	0.07	0.03	0.09
TLHGV	-0.00	-0.01	-0.00	-0.04	0.06	-0.00	-0.04	0.02	1.00	0.13	0.09	0.08	-0.03	0.09	0.07	0.18	0.11	0.08	0.07	0.10	0.03	0.06	0.06	0.03	0.05	0.02	0.13	0.01	0.16
Str	0.18	0.15	0.26	0.24	0.10	0.06	0.22	0.18	0.13	1.00	0.16	0.16	0.13	0.13	0.16	0.13	0.08	0.13	0.07	0.14	0.12	0.12	0.16	0.11	0.15	0.19	0.15	0.07	0.15
Kat	0.29	0.27	0.20	0.26	0.10	0.07	0.04	0.03	0.09	0.16	1.00	0.20	0.17	0.31	0.14	0.18	0.09	0.20	0.05	0.13	0.10	0.09	0.10	0.12	0.09	0.11	0.09	0.03	0.14
Typ	0.06	0.11	0.08	0.16	0.21	0.03	0.18	0.10	0.08	0.16	0.20	1.00	0.30	0.63	0.09	0.25	0.10	0.24	0.16	0.15	0.20	0.08	0.08	0.19	0.08	0.13	0.11	0.04	0.13
Betei	0.14	0.09	0.11	0.08	-0.13	-0.02	-0.07	0.02	-0.03	0.13	0.17	0.30	1.00	0.29	0.10	0.21	0.42	0.19	0.05	0.11	0.18	0.10	0.09	0.12	0.42	0.10	0.11	0.04	0.13
UArt1	0.16	0.20	0.20	0.27	0.24	0.07	0.19	0.15	0.09	0.13	0.31	0.63	0.29	1.00	0.16	0.21	0.11	0.29	0.16	0.22	0.19	0.11	0.12	0.19	0.09	0.15	0.14	0.08	0.12
UArt2	0.05	0.07	0.07	0.06	0.03	0.02	0.06	0.10	0.07	0.16	0.14	0.09	0.10	0.16	1.00	0.14	0.07	0.26	0.03	0.11	0.13	0.09	0.11	0.08	0.03	0.12	0.10	0.07	0.14
AUrs1	0.23	0.21	0.22	0.25	0.22	0.02	0.25	0.12	0.18	0.13	0.18	0.25	0.21	0.21	0.14	1.00	0.45	0.13	0.05	0.15	0.20	0.11	0.11	0.49	0.59	0.08	0.12	0.06	0.16
AUrs2	0.15	0.20	0.11	0.19	0.15	0.00	0.13	0.09	0.11	0.08	0.09	0.10	0.42	0.11	0.07	0.45	1.00	0.05	0.01	0.06	0.13	0.04	0.04	0.18	0.65	0.05	0.10	0.02	0.12
AufHi	0.16	0.14	0.13	0.09	0.14	0.03	0.14	0.07	0.08	0.13	0.20	0.24	0.19	0.29	0.26	0.13	0.05	1.00	0.03	0.12	0.18	0.08	0.09	0.15	0.03	0.12	0.11	0.05	0.12
Alkoh	0.00	0.02	-0.06	-0.03	0.07	-0.00	0.08	0.01	0.07	0.07	0.05	0.16	0.05	0.16	0.03	0.05	0.01	0.03	1.00	0.08	0.01	0.17	0.14	0.02	0.05	0.10	0.04	0.01	0.10
Char1	0.05	0.04	0.09	0.07	0.14	0.01	0.11	0.08	0.10	0.14	0.13	0.15	0.11	0.22	0.11	0.15	0.06	0.12	0.08	1.00	0.62	0.07	0.09	0.10	0.04	0.08	0.09	0.06	0.12
Char2	0.04	0.04	0.06	0.02	0.12	0.01	0.07	0.04	0.03	0.12	0.10	0.20	0.18	0.19	0.13	0.20	0.13	0.18	0.01	0.62	1.00	0.08	0.08	0.16	0.02	0.13	0.15	0.06	0.07
Lich1	0.02	0.04	0.13	0.03	0.18	0.02	0.12	0.04	0.06	0.12	0.09	0.08	0.10	0.11	0.09	0.11	0.04	0.08	0.17	0.07	0.08	1.00	0.71	0.43	0.05	0.08	0.11	0.06	0.20
Lich2	0.02	0.01	0.08	0.02	0.17	0.02	0.09	0.04	0.06	0.16	0.10	0.08	0.09	0.12	0.11	0.11	0.04	0.09	0.14	0.09	0.08	0.71	1.00	0.15	0.02	0.06	0.14	0.07	0.23
Zust1	0.10	0.08	0.10	0.09	0.13	0.02	0.18	0.06	0.03	0.11	0.12	0.19	0.12	0.19	0.08	0.49	0.18	0.15	0.02	0.10	0.16	0.43	0.15	1.00	0.16	0.09	0.09	0.05	0.27
Zust2	0.10	0.16	0.07	0.16	0.01	0.00	0.04	0.01	0.05	0.15	0.09	0.08	0.42	0.09	0.03	0.59	0.65	0.03	0.05	0.04	0.02	0.05	0.02	0.16	1.00	0.04	0.10	0.02	0.20
Fstf	0.03	0.03	0.03	0.04	0.05	0.07	0.03	0.03	0.02	0.19	0.11	0.13	0.10	0.15	0.12	0.08	0.05	0.12	0.10	0.08	0.13	0.08	0.06	0.09	0.04	1.00	0.10	0.06	0.13
WoTag	0.07	0.10	0.08	0.12	0.10	0.09	0.13	0.07	0.13	0.15	0.09	0.11	0.11	0.14	0.10	0.12	0.10	0.11	0.04	0.09	0.15	0.11	0.14	0.09	0.10	0.10	1.00	0.18	0.15
FeiTag	-0.02	-0.01	-0.02	0.00	0.00	-0.01	0.03	0.03	0.01	0.07	0.03	0.04	0.04	0.08	0.07	0.06	0.02	0.05	0.01	0.06	0.06	0.06	0.07	0.05	0.02	0.06	0.18	1.00	0.21
Month	0.12	0.11	0.14	0.13	0.11	0.11	0.16	0.09	0.16	0.15	0.14	0.13	0.13	0.12	0.14	0.16	0.12	0.12	0.10	0.12	0.07	0.20	0.23	0.27	0.20	0.13	0.15	0.21	1.00

Table A.27: Correlation matrix for BAYSIS selected data (Jam Initiator), calculated with Cramer's V, η, τ, r_{pq}, r

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	nan	0.000	0.000	0.000	0.002	0.361	0.000	0.671	0.921	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.965	0.000	0.000	0.000	0.000	0.000	0.000	0.325	0.000	0.576	0.000
TAvg	0.000	nan	0.000	0.000	0.011	0.335	0.021	0.680	0.724	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.527	0.000	0.000	0.000	0.000	0.000	0.000	0.206	0.000	0.716	0.000
SMax	0.000	0.000	nan	0.000	0.016	0.869	0.000	0.549	0.978	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.113	0.000	0.000	0.000	0.000	0.000	0.000	0.346	0.000	0.588	0.000
SAvg	0.000	0.000	0.000	nan	0.742	0.273	0.004	0.171	0.288	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.470	0.000	0.000	0.000	0.000	0.000	0.000	0.197	0.000	0.952	0.000
TDist	0.002	0.011	0.016	0.742	nan	0.449	0.000	0.827	0.110	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.043	0.000	0.000	0.000	0.000	0.000	0.000	0.094	0.000	0.970	0.000
SDist	0.361	0.335	0.869	0.273	0.449	nan	0.061	0.946	0.903	0.000	0.000	0.000	0.615	0.000	0.000	0.000	0.130	0.000	0.901	0.000	0.000	0.000	0.000	0.000	0.000	0.043	0.000	0.757	0.000
Cov	0.000	0.021	0.000	0.004	0.000	0.061	nan	0.494	0.245	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.222	0.000	0.336	0.000
TLCar	0.671	0.680	0.549	0.171	0.827	0.946	0.494	nan	0.595	0.000	0.000	0.000	0.400	0.000	0.000	0.000	0.000	0.000	0.808	0.000	0.000	0.000	0.000	0.000	0.000	0.307	0.000	0.360	0.000
TLHGV	0.921	0.724	0.978	0.288	0.110	0.903	0.245	0.595	nan	0.000	0.000	0.000	0.317	0.000	0.000	0.000	0.000	0.000	0.043	0.000	0.000	0.000	0.000	0.000	0.000	0.427	0.000	0.862	0.000
Str	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.040	0.022	0.812	0.808	0.015	0.714	1.000	0.612	0.995	0.229	0.699	0.748	0.106	0.914	0.187	0.000	0.113	0.995	0.013
Kat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	nan	0.000	0.000	0.000	0.004	0.000	0.534	0.000	0.516	0.000	0.037	0.049	0.028	0.000	0.124	0.157	0.705	0.842	0.039
Typ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.000	nan	0.000	0.000	0.621	0.000	0.169	0.000	0.002	0.000	0.000	0.591	0.440	0.000	0.433	0.002	0.076	0.931	0.240
Betei	0.000	0.001	0.000	0.002	0.000	0.615	0.009	0.400	0.317	0.812	0.000	0.000	nan	0.000	0.543	0.000	0.000	0.000	0.980	0.319	0.002	0.494	0.601	0.055	0.000	0.598	0.146	0.992	0.184
UArt1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.808	0.000	0.000	0.000	nan	0.000	0.000	0.282	0.000	0.018	0.000	0.001	0.288	0.206	0.000	0.756	0.000	0.002	0.819	0.379
UArt2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.004	0.621	0.543	0.000	nan	0.021	0.997	0.000	0.999	0.115	0.079	0.583	0.159	0.839	0.996	0.011	0.442	0.779	0.008
AUrs1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.714	0.000	0.000	0.000	0.000	0.021	nan	0.000	0.161	0.999	0.013	0.003	0.770	0.772	0.000	0.000	1.000	0.453	0.998	0.000
AUrs2	0.000	0.000	0.000	0.000	0.000	0.130	0.000	0.000	0.000	1.000	0.534	0.169	0.000	0.282	0.997	0.000	nan	1.000	1.000	0.996	0.059	1.000	0.997	0.000	0.000	1.000	0.302	1.000	0.379
AufHi	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.612	0.000	0.000	0.000	0.000	0.000	0.161	1.000	nan	0.999	0.036	0.002	0.910	0.772	0.002	0.999	0.020	0.111	0.976	0.514
Alkoh	0.965	0.527	0.113	0.470	0.043	0.901	0.024	0.808	0.043	0.995	0.516	0.002	0.980	0.018	0.999	0.999	1.000	0.999	nan	0.296	0.887	0.000	0.001	0.961	0.198	0.304	0.994	0.779	0.720
Char1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.229	0.000	0.000	0.319	0.000	0.115	0.013	0.996	0.036	0.296	nan	0.000	0.384	0.141	0.026	0.908	0.907	0.543	0.654	0.512
Char2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.699	0.037	0.000	0.002	0.001	0.079	0.003	0.059	0.002	0.887	0.000	nan	0.150	0.079	0.000	0.633	0.084	0.011	0.080	0.969
Lich1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.748	0.049	0.591	0.494	0.288	0.583	0.770	1.000	0.910	0.000	0.384	0.150	nan	0.000	0.000	0.560	0.809	0.214	0.486	0.000
Lich2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.106	0.028	0.440	0.601	0.206	0.159	0.772	0.997	0.772	0.001	0.141	0.079	0.000	nan	0.000	0.797	0.966	0.004	0.115	0.000
Zust1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.914	0.000	0.000	0.055	0.000	0.839	0.000	0.000	0.002	0.961	0.026	0.000	0.000	0.000	nan	0.000	0.703	0.603	0.588	0.000
Zust2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.187	0.124	0.433	0.000	0.756	0.996	0.000	0.000	0.999	0.198	0.908	0.633	0.560	0.797	0.000	nan	0.990	0.417	0.534	0.001
Fstf	0.325	0.206	0.346	0.197	0.094	0.043	0.222	0.307	0.427	0.000	0.157	0.002	0.598	0.000	0.011	1.000	1.000	0.020	0.304	0.907	0.084	0.809	0.966	0.703	0.990	nan	0.414	0.875	0.165
WoTag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.113	0.705	0.076	0.146	0.002	0.442	0.453	0.302	0.111	0.994	0.543	0.011	0.214	0.004	0.603	0.417	0.414	nan	0.001	0.002
FeiTag	0.576	0.716	0.588	0.952	0.970	0.757	0.336	0.360	0.862	0.995	0.842	0.931	0.992	0.819	0.779	0.998	1.000	0.976	0.779	0.654	0.080	0.486	0.115	0.588	0.534	0.875	0.001	nan	0.000
Month	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.039	0.240	0.184	0.379	0.008	0.000	0.379	0.514	0.720	0.512	0.969	0.000	0.000	0.000	0.001	0.165	0.002	0.000	nan

Table A.28: Significancy matrix for BAYSIS selected data (Jam Initiator)

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	NaN	r	r	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TAvg	r	NaN	r	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SMax	r	r	NaN	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SAvg	r	r	r	NaN	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TDist	r	r	r	r	NaN	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SDist	r	r	r	r	r	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Cov	r	r	r	r	r	r	NaN	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLCar	r	r	r	r	r	r	r	NaN	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLHGV	r	r	r	r	r	r	r	r	NaN	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Str	η	η	η	η	η	η	η	η	η	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Kat	η	η	η	η	η	η	η	η	η	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Typ	η	η	η	η	η	η	η	η	η	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Betei	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt1	η	η	η	η	η	η	η	η	η	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V
AufHi	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V
Alkoh	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V
Char1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V
Char2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V
Lich1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V
Lich2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V
Zust1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V
Zust2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V
Fstf	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V
WoTag	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V
FeiTag	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V
Month	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN

Table A.29: Coefficient matrix for BAYSIS selected data (Jam Initiator)

	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A6	1.00												
A9	1.00	1.00											
A70	1.00	1.00	1.00										
A99	0.85	1.00	1.00	1.00									
A93	1.00	1.00	1.00	1.00	1.00								
A94	0.01	1.00	0.06	1.00	0.28	1.00							
A7	.00	1.00	1.00	1.00	1.00	1.00	1.00						
A73	0.00	1.00	0.00	1.00	0.23	1.00	1.00	1.00					
A96	0.00	1.00	0.35	1.00	1.00	1.00	1.00	1.00	1.00				
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A92	0.01	1.00	0.24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table A.30: Pairwise Wilcoxon T-test for *Street* and *Maximal Temporal Extent* (Jam Effector) complete

	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A6	1.00												
A9	1.00	1.00											
A70	1.00	1.00	1.00										
A99	0.02	1.00	1.00	1.00									
A93	1.00	1.00	1.00	1.00	1.00								
A94	0.11	1.00	0.53	1.00	1.00	1.00							
A7	1.00	1.00	1.00	1.00	1.00	1.00	0.61						
A73	0.00	1.00	0.00	1.00	1.00	1.00	1.00	0.02					
A96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87				
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.31: Pairwise Wilcoxon T-test for *Strasse* and *Average Temporal Extent* (Jam Effector) complete

	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A9	0.00	1.00											
A93	0.07	0.13	1.00	1.00	1.00								
A94	0.01	0.03	0.41	1.00	0.24	1.00							
A7	0.08	0.65	1.00	1.00	1.00	1.00	1.00						
A73	0.00	0.00	0.00	1.00	0.00	1.00	1.00	0.95					
A96	0.13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.18				
A92	0.00	0.00	0.04	1.00	0.04	1.00	1.00	1.00	1.00	1.00	1.00		

Table A.32: Pairwise Wilcoxon T-test for *Street* and *Maximal Spatial Extent* (Jam Effector) complete

	A3	A6	A9	A70	A99	A93	A94	A7	A73	A96	A995	A92	A95
A6	1.00												
A9	0.01	1.00											
A70	1.00	1.00	1.00										
A99	1.00	1.00	0.00	1.00									
A93	1.00	1.00	1.00	1.00	1.00								
A94	1.00	1.00	1.00	1.00	1.00	1.00							
A7	0.25	1.00	1.00	1.00	0.02	1.00	1.00						
A73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A96	0.21	1.00	1.00	1.00	0.02	1.00	1.00	1.00	1.00				
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A92	0.03	0.43	1.00	1.00	0.01	1.00	1.00	1.00	0.61	1.00	1.00		
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.33: Pairwise Wilcoxon T-test for *Street* and *Coverage* (Jam Effector) complete

A.1 BAYSIS Selected Data - Jam Follower

A.2 BAYSIS Predictability

	Di	Do	Fr	Mi	Mo	Sa
Do	1.00					
Fr	1.00	1.00				
Mi	1.00	1.00	1.00			
Mo	1.00	1.00	1.00	1.00		
Sa	1.00	1.00	1.00	1.00	1.00	
So	1.00	1.00	1.00	1.00	1.00	1.00

Table A.34: Pairwise Wilcoxon T-test for *WoTag* and *Maximal Spatial Extent* (Jam Effector)

	Di	Do	Fr	Mi	Mo	Sa
Do	1.00					
Fr	1.00	1.00				
Mi	0.03	1.00	1.00			
Mo	0.82	1.00	1.00	1.00		
Sa	1.00	1.00	1.00	0.82	1.00	
So	1.00	1.00	1.00	0.05	0.65	1.00

Table A.35: Pairwise Wilcoxon T-test for *WoTag* and *Time-loss HGV* (Jam Effector) complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	1.00	1.00									
Apr	1.00	1.00	1.00								
May	1.00	1.00	1.00	1.00							
Jun	1.00	1.00	1.00	1.00	1.00						
Jul	1.00	1.00	0.53	1.00	1.00	1.00					
Aug	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
Sep	1.00	1.00	0.82	1.00	1.00	1.00	1.00	1.00			
Oct	1.00	1.00	1.00	1.00	1.00	1.00	0.17	1.00	0.31		
Nov	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.15	0.01	1.00	
Dec	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93

Table A.36: Pairwise Wilcoxon T-test for *Month* and *Maximal Spatial Extent* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	1.00	1.00									
Apr	1.00	1.00	1.00								
May	1.00	1.00	1.00	1.00							
Jun	1.00	1.00	1.00	1.00	1.00						
Jul	1.00	1.00	1.00	1.00	1.00	1.00					
Aug	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
Sep	1.00	1.00	0.79	1.00	1.00	1.00	1.00	1.00			
Oct	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
Nov	1.00	1.00	1.00	1.00	1.00	1.00	0.83	1.00	0.08	1.00	
Dec	1.00	1.00	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.07

Table A.37: Pairwise Wilcoxon T-test for *Month* and *Average Spatial Extent* (Jam Effector) complete

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	1.00	0.80	0.50	0.43	0.00	0.00	-0.19	0.03	-0.03	0.28	0.15	0.05	0.07	0.06	0.09	0.12	0.02	0.16	-0.02	0.09	0.05	0.05	0.05	0.11	0.03	0.00	0.12	0.03	0.19
TAvg	0.80	1.00	0.22	0.46	0.00	0.00	0.18	0.03	-0.02	0.24	0.16	0.08	0.09	0.08	0.08	0.09	0.01	0.24	0.02	0.03	0.04	0.05	0.02	0.08	0.00	0.02	0.15	0.02	0.18
SMax	0.50	0.22	1.00	0.62	0.00	0.00	-0.48	0.00	-0.09	0.36	0.16	0.05	0.05	0.12	0.08	0.08	0.02	0.09	-0.05	0.10	0.02	0.06	0.05	0.08	0.05	0.06	0.11	0.07	0.19
SAvg	0.43	0.46	0.62	1.00	0.00	0.00	0.18	-0.03	-0.10	0.31	0.24	0.04	0.10	0.17	0.10	0.13	0.03	0.04	-0.03	0.06	0.00	0.01	0.02	0.08	0.11	0.10	0.11	0.07	0.17
TDist	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SDist	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cov	-0.19	0.18	-0.48	0.18	0.00	0.00	1.00	-0.05	0.04	0.31	0.07	0.11	0.05	0.11	0.14	0.12	0.06	0.11	0.05	0.12	0.03	0.13	0.10	0.11	0.00	0.06	0.20	-0.01	0.19
TLCar	0.03	0.03	0.00	-0.03	0.00	0.00	-0.05	1.00	0.00	0.12	0.01	0.11	0.00	0.11	0.10	0.12	0.03	0.05	-0.08	0.09	0.01	0.02	0.02	0.07	0.03	-0.05	0.08	-0.04	0.10
TLHGV	-0.03	-0.02	-0.09	-0.10	0.00	0.00	0.04	0.00	1.00	0.21	0.13	0.10	0.01	0.08	0.08	0.14	0.07	0.05	-0.05	0.12	0.01	0.03	0.03	0.11	0.05	-0.01	0.14	0.05	0.21
Str	0.28	0.24	0.36	0.31	0.00	0.00	0.31	0.12	0.21	1.00	0.14	0.15	0.12	0.20	0.15	0.14	0.11	0.12	0.08	0.22	0.15	0.18	0.16	0.22	0.14	0.18	0.16	0.12	0.15
Kat	0.15	0.16	0.16	0.24	0.00	0.00	0.07	0.01	0.13	0.14	1.00	0.22	0.27	0.38	0.11	0.07	0.04	0.06	0.12	0.08	0.05	0.07	0.06	0.05	0.07	0.10	0.13	0.08	0.12
Typ	0.05	0.08	0.05	0.04	0.00	0.00	0.11	0.11	0.10	0.15	0.22	1.00	0.32	0.51	0.10	0.27	0.08	0.25	0.07	0.14	0.13	0.12	0.13	0.16	0.26	0.14	0.13	0.08	0.15
Betei	0.07	0.09	0.05	0.10	0.00	0.00	0.05	0.00	0.01	0.12	0.27	0.32	1.00	0.34	0.10	0.15	0.02	0.29	0.10	0.09	0.10	0.10	0.06	0.09	0.22	0.12	0.12	0.12	0.15
UArt1	0.06	0.08	0.12	0.17	0.00	0.00	0.11	0.11	0.08	0.20	0.38	0.51	0.34	1.00	0.17	0.24	0.11	0.39	0.09	0.16	0.15	0.08	0.07	0.15	0.22	0.16	0.13	0.09	0.13
UArt2	0.09	0.08	0.08	0.10	0.00	0.00	0.14	0.10	0.08	0.15	0.11	0.10	0.10	0.17	1.00	0.10	0.01	0.39	0.03	0.09	0.12	0.09	0.07	0.10	0.10	0.10	0.10	0.03	0.11
AUrs1	0.12	0.09	0.08	0.13	0.00	0.00	0.12	0.12	0.14	0.14	0.07	0.27	0.15	0.24	0.10	1.00	0.28	0.30	0.03	0.07	0.02	0.10	0.08	0.42	0.52	0.05	0.10	0.06	0.18
AUrs2	0.02	0.01	0.02	0.03	0.00	0.00	0.06	0.03	0.07	0.11	0.04	0.08	0.02	0.11	0.01	0.28	1.00	0.01	0.01	0.01	0.01	0.12	0.06	0.29	0.01	0.05	0.10	0.01	0.15
AufHi	0.16	0.24	0.09	0.04	0.00	0.00	0.11	0.05	0.05	0.12	0.06	0.25	0.29	0.39	0.39	0.30	0.01	1.00	0.02	0.05	0.15	0.05	0.06	0.12	0.15	0.11	0.11	0.02	0.13
Alkoh	-0.02	0.02	-0.05	-0.03	0.00	0.00	0.05	-0.08	-0.05	0.08	0.12	0.07	0.10	0.09	0.03	0.03	0.01	0.02	1.00	0.09	0.05	0.14	0.14	0.08	0.05	0.04	0.08	0.02	0.09
Char1	0.09	0.03	0.10	0.06	0.00	0.00	0.12	0.09	0.12	0.22	0.08	0.14	0.09	0.16	0.09	0.07	0.01	0.05	0.09	1.00	0.48	0.07	0.04	0.05	0.08	0.10	0.10	0.04	0.12
Char2	0.05	0.04	0.02	0.00	0.00	0.00	0.03	0.01	0.01	0.15	0.05	0.13	0.10	0.15	0.12	0.02	0.01	0.15	0.05	0.48	1.00	0.06	0.04	0.03	0.06	0.08	0.10	0.04	0.12
Lich1	0.05	0.05	0.06	0.01	0.00	0.00	0.13	0.02	0.03	0.18	0.07	0.12	0.10	0.08	0.09	0.10	0.12	0.05	0.14	0.07	0.06	1.00	0.71	0.13	0.05	0.09	0.12	0.03	0.32
Lich2	0.05	0.02	0.05	0.02	0.00	0.00	0.10	0.02	0.03	0.16	0.06	0.13	0.06	0.07	0.07	0.08	0.06	0.06	0.14	0.04	0.04	0.71	1.00	0.13	0.04	0.09	0.11	0.01	0.32
Zust1	0.11	0.08	0.08	0.08	0.00	0.00	0.11	0.07	0.11	0.22	0.05	0.16	0.09	0.15	0.10	0.42	0.29	0.12	0.08	0.05	0.03	0.13	0.13	1.00	0.22	0.07	0.11	0.05	0.27
Zust2	0.03	0.00	0.05	0.11	0.00	0.00	0.00	0.03	0.05	0.14	0.07	0.26	0.22	0.22	0.10	0.52	0.01	0.15	0.05	0.08	0.06	0.05	0.04	0.22	1.00	0.08	0.10	0.04	0.25
Fstf	0.00	0.02	0.06	0.10	0.00	0.00	0.06	-0.05	-0.01	0.18	0.10	0.14	0.12	0.16	0.10	0.05	0.05	0.11	0.04	0.10	0.08	0.09	0.09	0.07	0.08	1.00	0.10	0.13	0.13
WoTag	0.12	0.15	0.11	0.11	0.00	0.00	0.20	0.08	0.14	0.16	0.13	0.13	0.12	0.13	0.10	0.10	0.10	0.11	0.08	0.10	0.10	0.12	0.11	0.11	0.10	0.10	1.00	0.14	0.16
FeiTag	0.03	0.02	0.07	0.07	0.00	0.00	-0.01	-0.04	0.05	0.12	0.08	0.08	0.12	0.09	0.03	0.06	0.01	0.02	0.02	0.04	0.04	0.03	0.01	0.05	0.04	0.13	0.14	1.00	0.16
Month	0.19	0.18	0.19	0.17	0.00	0.00	0.19	0.10	0.21	0.15	0.12	0.15	0.15	0.13	0.11	0.18	0.15	0.13	0.09	0.12	0.12	0.32	0.32	0.27	0.25	0.13	0.16	0.16	1.00

Table A.38: Correlation matrix for BAYSIS selected data (Jam Effector), calculated with Cramer's V, η, τ, r_{pq}, r

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	nan	0.000	0.000	0.000	nan	nan	0.000	0.363	0.495	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.577	0.000	0.000	0.000	0.000	0.000	0.000	0.979	0.000	0.345	0.000
TAvg	0.000	nan	0.000	0.000	nan	nan	0.000	0.494	0.560	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.595	0.000	0.000	0.000	0.000	0.000	0.000	0.489	0.000	0.577	0.000
SMax	0.000	0.000	nan	0.000	nan	nan	0.000	0.960	0.014	0.000	0.000	0.000	0.077	0.000	0.000	0.000	0.000	0.000	0.170	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.000	0.030	0.000
SAvg	0.000	0.000	0.000	nan	nan	nan	0.000	0.436	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.423	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.023	0.000
TDist	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
SDist	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
Cov	0.000	0.000	0.000	0.000	nan	nan	nan	0.194	0.263	0.000	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.135	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.000	0.777	0.000
TLCar	0.363	0.494	0.960	0.436	nan	nan	0.194	nan	1.000	0.000	0.000	0.000	0.875	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.078	0.000	0.242	0.000
TLHGV	0.495	0.560	0.014	0.005	nan	nan	0.263	1.000	nan	0.000	0.000	0.000	0.690	0.000	0.000	0.000	0.000	0.000	0.201	0.000	0.000	0.000	0.000	0.000	0.000	0.700	0.000	0.119	0.000
Str	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	nan	0.265	0.102	0.790	0.000	0.067	0.163	0.840	0.838	0.989	0.000	0.203	0.005	0.041	0.000	0.378	0.000	0.001	0.660	0.024
Kat	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.265	nan	0.000	0.000	0.000	0.064	0.934	0.847	0.808	0.012	0.232	0.670	0.321	0.467	0.828	0.291	0.306	0.013	0.161	0.537
Typ	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.102	0.000	nan	0.000	0.000	0.196	0.000	0.227	0.000	0.478	0.000	0.012	0.008	0.003	0.000	0.000	0.001	0.008	0.323	0.031
Betei	0.011	0.003	0.077	0.000	nan	nan	0.062	0.875	0.690	0.790	0.000	0.000	nan	0.000	0.178	0.000	1.000	0.000	0.268	0.531	0.303	0.339	0.915	0.477	0.000	0.024	0.033	0.081	0.005
UArt1	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.361	0.000	0.589	0.000	0.031	0.832	0.956	0.002	0.000	0.000	0.003	0.576	0.276
UArt2	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.067	0.064	0.196	0.178	0.000	nan	0.223	1.000	0.000	0.994	0.433	0.109	0.442	0.879	0.164	0.228	0.541	0.234	0.996	0.900
AUrs1	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.163	0.934	0.000	0.000	0.000	0.223	nan	0.000	0.000	0.999	0.964	1.000	0.374	0.830	0.000	0.000	1.000	0.210	0.895	0.000
AUrs2	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.840	0.847	0.227	1.000	0.361	1.000	0.000	nan	1.000	0.985	1.000	0.991	0.000	0.199	0.000	0.991	0.997	0.398	0.981	0.075
AufHi	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.838	0.808	0.000	0.000	0.000	0.000	0.000	1.000	nan	0.995	0.930	0.003	0.934	0.782	0.002	0.003	0.219	0.077	0.980	0.148
Alkoh	0.577	0.595	0.170	0.423	nan	nan	0.135	0.028	0.201	0.989	0.012	0.478	0.268	0.589	0.994	0.999	0.985	0.995	nan	0.240	0.213	0.001	0.001	0.236	0.213	0.989	0.738	0.514	0.835
Char1	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.000	0.232	0.000	0.531	0.000	0.433	0.964	1.000	0.930	0.240	nan	0.000	0.488	0.946	0.921	0.353	0.476	0.259	0.918	0.502
Char2	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.203	0.670	0.012	0.303	0.031	0.109	1.000	0.991	0.003	0.213	0.000	nan	0.306	0.565	0.843	0.088	0.648	0.423	0.304	0.482
Lich1	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.005	0.321	0.008	0.339	0.832	0.442	0.374	0.000	0.934	0.001	0.488	0.306	nan	0.000	0.000	0.438	0.544	0.065	0.677	0.000
Lich2	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.041	0.467	0.003	0.915	0.956	0.879	0.830	0.199	0.782	0.001	0.946	0.565	0.000	nan	0.000	0.565	0.669	0.175	0.923	0.000
Zust1	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.000	0.828	0.000	0.477	0.002	0.164	0.000	0.000	0.002	0.236	0.921	0.843	0.000	0.000	nan	0.000	0.956	0.139	0.643	0.000
Zust2	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.378	0.291	0.000	0.000	0.000	0.228	0.000	0.991	0.003	0.213	0.353	0.088	0.438	0.565	0.000	nan	0.752	0.410	0.304	0.000
Fstf	0.979	0.489	0.021	0.001	nan	nan	0.038	0.078	0.700	0.000	0.306	0.001	0.024	0.000	0.541	1.000	0.997	0.219	0.989	0.476	0.648	0.544	0.669	0.956	0.752	nan	0.241	0.098	0.277
WoTag	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.001	0.013	0.008	0.033	0.003	0.234	0.210	0.398	0.077	0.738	0.259	0.423	0.065	0.175	0.139	0.410	0.241	nan	0.044	0.000
FeiTag	0.345	0.577	0.030	0.023	nan	nan	0.777	0.242	0.119	0.660	0.161	0.323	0.081	0.576	0.996	0.895	0.981	0.980	0.514	0.918	0.304	0.677	0.923	0.643	0.304	0.098	0.044	nan	0.071
Month	0.000	0.000	0.000	0.000	nan	nan	0.000	0.000	0.000	0.024	0.537	0.031	0.005	0.276	0.900	0.000	0.075	0.148	0.835	0.502	0.482	0.000	0.000	0.000	0.000	0.277	0.000	0.071	nan

Table A.39: Significancy matrix for BAYSIS selected data (Jam Effector)

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	NaN	r	r	r	NaN	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TAvg	r	NaN	r	r	NaN	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SMax	r	r	NaN	r	NaN	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SAvg	r	r	r	NaN	NaN	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TDist	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SDist	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Cov	r	r	r	r	NaN	NaN	NaN	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLCar	r	r	r	r	NaN	NaN	r	NaN	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLHGV	r	r	r	r	NaN	NaN	r	r	NaN	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Str	η	η	η	η	NaN	NaN	η	η	η	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Kat	η	η	η	η	NaN	NaN	η	η	η	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Typ	η	η	η	η	NaN	NaN	η	η	η	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Betei	τ	τ	τ	τ	NaN	NaN	τ	τ	τ	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt1	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt2	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs1	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs2	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V
AufHi	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V
Alkoh	r_{pq}	r_{pq}	r_{pq}	r_{pq}	NaN	NaN	r_{pq}	r_{pq}	r_{pq}	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V
Char1	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V
Char2	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V
Lich1	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V
Lich2	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V
Zust1	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V
Zust2	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V
Fstf	τ	τ	τ	τ	NaN	NaN	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V
WoTag	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V
FeiTag	τ	τ	τ	τ	NaN	NaN	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V
Month	η	η	η	η	NaN	NaN	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN

Table A.40: Coefficient matrix for BAYSIS selected data (Jam Effector)

	0	72	73	80	82	88
72	1.00					
73	0.10	0.35				
80	1.00		1.00			
82	1.00		1.00			
88	0.40	0.04	1.00	1.00	0.55	
89	1.00		1.00			1.00

Table A.41: Pairwise Wilcoxon T-test for *AUrs1* and *Spatial Distance* (Jam Follower)

	0	72	73	80	82	88
72	0.28					
73	1.00	1.00				
80	1.00	1.00	1.00			
82	1.00	1.00	1.00	1.00		
88	0.33	0.79	0.56	1.00	1.00	
89	1.00	1.00	1.00	1.00	1.00	1.00

Table A.42: Pairwise Wilcoxon T-test for *AUrs1* and *Coverage* (Jam Follower)

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	1.00	0.81	0.49	0.50	-0.24	-0.05	-0.12	0.05	-0.02	0.29	0.15	0.07	0.07	0.10	0.12	0.12	0.02	0.22	0.04	0.09	0.04	0.02	0.02	0.13	0.05	-0.02	0.11	-0.00	0.19
TAvg	0.81	1.00	0.19	0.45	-0.17	-0.03	0.24	0.01	0.01	0.22	0.13	0.12	0.08	0.16	0.07	0.07	0.01	0.33	0.04	0.05	0.03	0.10	0.08	0.09	0.04	-0.02	0.15	0.00	0.15
SMax	0.49	0.19	1.00	0.66	-0.21	-0.04	-0.50	0.05	-0.10	0.35	0.18	0.10	0.07	0.17	0.09	0.10	0.02	0.11	0.03	0.07	0.04	0.08	0.06	0.09	0.04	0.05	0.15	0.05	0.21
SAvg	0.50	0.45	0.66	1.00	-0.24	-0.11	0.09	0.01	-0.09	0.32	0.26	0.06	0.10	0.19	0.12	0.10	0.05	0.08	0.05	0.10	0.09	0.02	0.03	0.09	0.02	0.03	0.16	0.06	0.15
TDist	-0.24	-0.17	-0.21	-0.24	1.00	0.08	0.03	-0.04	0.04	0.19	0.12	0.13	-0.09	0.24	0.15	0.20	0.02	0.07	0.02	0.09	0.08	0.01	0.03	0.10	0.05	0.01	0.15	-0.02	0.09
SDist	-0.05	-0.03	-0.04	-0.11	0.08	1.00	0.00	-0.01	0.09	0.23	0.04	0.12	-0.03	0.12	0.10	0.17	0.02	0.07	-0.04	0.05	0.00	0.01	0.02	0.05	0.02	0.05	0.13	0.06	0.13
Cov	-0.12	0.24	-0.50	0.09	0.03	0.00	1.00	-0.04	0.12	0.35	0.06	0.15	0.03	0.22	0.14	0.22	0.06	0.17	-0.01	0.08	0.05	0.15	0.14	0.14	0.02	-0.01	0.18	0.01	0.23
TLCar	0.05	0.01	0.05	0.01	-0.04	-0.01	-0.04	1.00	-0.01	0.09	0.07	0.03	0.02	0.13	0.11	0.07	0.04	0.09	-0.02	0.12	0.01	0.09	0.09	0.07	0.04	-0.01	0.14	0.03	0.11
TLHGV	-0.02	0.01	-0.10	-0.09	0.04	0.09	0.12	-0.01	1.00	0.24	0.06	0.11	-0.09	0.17	0.12	0.23	0.10	0.09	-0.05	0.11	0.05	0.06	0.05	0.12	0.00	-0.01	0.20	0.03	0.18
Str	0.29	0.22	0.35	0.32	0.19	0.23	0.35	0.09	0.24	1.00	0.15	0.23	0.15	0.20	0.14	0.24	0.14	0.17	0.12	0.23	0.41	0.23	0.17	0.25	0.16	0.20	0.18	0.13	0.19
Kat	0.15	0.13	0.18	0.26	0.12	0.04	0.06	0.07	0.06	0.15	1.00	0.19	0.27	0.41	0.17	0.10	0.05	0.11	0.23	0.11	0.11	0.09	0.09	0.04	0.05	0.13	0.11	0.05	0.15
Typ	0.07	0.12	0.10	0.06	0.13	0.12	0.15	0.03	0.11	0.23	0.19	1.00	0.36	0.52	0.13	0.30	0.09	0.32	0.07	0.10	0.25	0.13	0.24	0.20	0.13	0.24	0.14	0.10	0.18
Betei	0.07	0.08	0.07	0.10	-0.09	-0.03	0.03	0.02	-0.09	0.15	0.27	0.36	1.00	0.33	0.20	0.20	0.03	0.37	0.06	0.11	0.26	0.13	0.11	0.17	0.07	0.17	0.12	0.07	0.17
UArt1	0.10	0.16	0.17	0.19	0.24	0.12	0.22	0.13	0.17	0.20	0.41	0.52	0.33	1.00	0.33	0.32	0.15	0.51	0.13	0.13	0.23	0.14	0.15	0.19	0.12	0.21	0.14	0.11	0.17
UArt2	0.12	0.07	0.09	0.12	0.15	0.10	0.14	0.11	0.12	0.14	0.17	0.13	0.20	0.33	1.00	0.12	0.01	0.42	0.03	0.08	0.13	0.09	0.07	0.11	0.03	0.12	0.13	0.06	0.16
AUrs1	0.12	0.07	0.10	0.10	0.20	0.17	0.22	0.07	0.23	0.24	0.10	0.30	0.20	0.32	0.12	1.00	0.33	0.35	0.03	0.11	0.15	0.21	0.19	0.57	0.16	0.08	0.13	0.16	0.21
AUrs2	0.02	0.01	0.02	0.05	0.02	0.02	0.06	0.04	0.10	0.14	0.05	0.09	0.03	0.15	0.01	0.33	1.00	0.02	0.01	0.01	0.01	0.16	0.08	0.33	0.01	0.06	0.11	0.01	0.18
AufHi	0.22	0.33	0.11	0.08	0.07	0.07	0.17	0.09	0.09	0.17	0.11	0.32	0.37	0.51	0.42	0.35	0.02	1.00	0.05	0.12	0.30	0.07	0.08	0.16	0.04	0.18	0.14	0.18	0.16
Alkoh	0.04	0.04	0.03	0.05	0.02	-0.04	-0.01	-0.02	-0.05	0.12	0.23	0.07	0.06	0.13	0.03	0.03	0.01	0.05	1.00	0.03	0.07	0.08	0.01	0.06	0.09	0.09	0.07	0.05	0.13
Char1	0.09	0.05	0.07	0.10	0.09	0.05	0.08	0.12	0.11	0.23	0.11	0.10	0.11	0.13	0.08	0.11	0.01	0.12	0.03	1.00	0.53	0.10	0.04	0.08	0.14	0.12	0.12	0.03	0.15
Char2	0.04	0.03	0.04	0.09	0.08	0.00	0.05	0.01	0.05	0.41	0.11	0.25	0.26	0.23	0.13	0.15	0.01	0.30	0.07	0.53	1.00	0.08	0.06	0.13	0.09	0.21	0.07	0.05	0.15
Lich1	0.02	0.10	0.08	0.02	0.01	0.01	0.15	0.09	0.06	0.23	0.09	0.13	0.13	0.14	0.09	0.21	0.16	0.07	0.08	0.10	0.08	1.00	0.71	0.19	0.10	0.11	0.22	0.05	0.34
Lich2	0.02	0.08	0.06	0.03	0.03	0.02	0.14	0.09	0.05	0.17	0.09	0.24	0.11	0.15	0.07	0.19	0.08	0.08	0.01	0.04	0.06	0.71	1.00	0.19	0.03	0.12	0.19	0.03	0.33
Zust1	0.13	0.09	0.09	0.09	0.10	0.05	0.14	0.07	0.12	0.25	0.04	0.20	0.17	0.19	0.11	0.57	0.33	0.16	0.06	0.08	0.13	0.19	0.19	1.00	0.19	0.08	0.17	0.10	0.30
Zust2	0.05	0.04	0.04	0.02	0.05	0.02	0.02	0.04	0.00	0.16	0.05	0.13	0.07	0.12	0.03	0.16	0.01	0.04	0.09	0.14	0.09	0.10	0.03	0.19	1.00	0.10	0.08	0.07	0.25
Fstf	-0.02	-0.02	0.05	0.03	0.01	0.05	-0.01	-0.01	-0.01	0.20	0.13	0.24	0.17	0.21	0.12	0.08	0.06	0.18	0.09	0.12	0.21	0.11	0.12	0.08	0.10	1.00	0.11	0.13	0.14
WoTag	0.11	0.15	0.15	0.16	0.15	0.13	0.18	0.14	0.20	0.18	0.11	0.14	0.12	0.14	0.13	0.13	0.11	0.14	0.07	0.12	0.07	0.22	0.19	0.17	0.08	0.11	1.00	0.16	0.18
FeiTag	-0.00	0.00	0.05	0.06	-0.02	0.06	0.01	0.03	0.03	0.13	0.05	0.10	0.07	0.11	0.06	0.16	0.01	0.18	0.05	0.03	0.05	0.05	0.03	0.10	0.07	0.13	0.16	1.00	0.19
Month	0.19	0.15	0.21	0.15	0.09	0.13	0.23	0.11	0.18	0.19	0.15	0.18	0.17	0.17	0.16	0.21	0.18	0.16	0.13	0.15	0.15	0.34	0.33	0.30	0.25	0.14	0.18	0.19	1.00

Table A.43: Correlation matrix for BAYSIS selected data (Jam Follower), calculated with Cramer's V, η, τ, r_{pq}, r

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	nan	0.000	0.000	0.000	0.000	0.277	0.010	0.320	0.738	0.000	0.000	0.000	0.049	0.000	0.000	0.000	0.000	0.000	0.416	0.000	0.000	0.000	0.000	0.000	0.000	0.502	0.000	0.930	0.000
TAvg	0.000	nan	0.000	0.000	0.000	0.566	0.000	0.798	0.789	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.340	0.000	0.000	0.000	0.000	0.000	0.000	0.499	0.000	0.904	0.000
SMax	0.000	0.000	nan	0.000	0.000	0.378	0.000	0.314	0.038	0.000	0.000	0.000	0.057	0.000	0.000	0.000	0.000	0.000	0.514	0.000	0.000	0.000	0.000	0.000	0.000	0.126	0.000	0.189	0.000
SAvg	0.000	0.000	0.000	nan	0.000	0.021	0.046	0.810	0.061	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.321	0.000	0.000	0.000	0.000	0.000	0.000	0.342	0.000	0.125	0.000
TDist	0.000	0.000	0.000	0.000	nan	0.094	0.464	0.343	0.365	0.000	0.000	0.000	0.044	0.000	0.000	0.001	0.000	0.000	0.620	0.000	0.000	0.169	0.000	0.491	0.000	0.842	0.000	0.584	0.000
SDist	0.277	0.566	0.378	0.021	0.094	nan	0.971	0.870	0.043	0.000	0.000	0.000	0.517	0.000	0.000	0.000	0.000	0.000	0.392	0.000	0.000	0.001	0.000	0.191	0.000	0.226	0.000	0.195	0.000
Cov	0.010	0.000	0.000	0.046	0.464	0.971	nan	0.337	0.008	0.000	0.000	0.000	0.453	0.000	0.000	0.000	0.000	0.000	0.769	0.000	0.000	0.000	0.000	0.000	0.000	0.840	0.000	0.800	0.000
TLCar	0.320	0.798	0.314	0.810	0.343	0.870	0.337	nan	0.803	0.000	0.000	0.000	0.639	0.000	0.000	0.000	0.000	0.000	0.604	0.000	0.000	0.000	0.000	0.000	0.000	0.822	0.000	0.360	0.000
TLHGV	0.738	0.789	0.038	0.061	0.365	0.043	0.008	0.803	nan	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.300	0.000	0.000	0.000	0.000	0.000	0.000	0.886	0.000	0.419	0.000
Str	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.507	0.000	0.696	0.000	0.772	0.000	0.616	0.128	0.817	0.000	0.000	0.001	0.207	0.000	0.390	0.000	0.037	0.738	0.001
Kat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.507	nan	0.000	0.000	0.000	0.001	0.729	0.867	0.043	0.000	0.068	0.111	0.287	0.252	0.980	0.793	0.134	0.631	0.727	0.593
Typ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.194	0.000	0.455	0.000	0.731	0.365	0.000	0.068	0.000	0.000	0.099	0.000	0.107	0.377	0.067
Betei	0.049	0.033	0.057	0.005	0.044	0.517	0.453	0.639	0.017	0.696	0.000	0.000	nan	0.000	0.000	0.000	1.000	0.000	0.975	0.770	0.000	0.290	0.588	0.007	0.940	0.001	0.453	0.932	0.055
UArt1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.164	0.000	0.465	0.544	0.002	0.339	0.188	0.002	0.533	0.000	0.344	0.677	0.045
UArt2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.772	0.001	0.194	0.000	0.000	nan	0.298	1.000	0.000	0.997	0.961	0.215	0.802	0.957	0.606	0.999	0.379	0.306	0.963	0.380
AUrs1	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.729	0.000	0.000	0.000	0.298	nan	0.000	0.000	0.998	0.525	0.107	0.000	0.001	0.000	0.076	0.988	0.312	0.055	0.000
AUrs2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.616	0.867	0.455	1.000	0.164	1.000	0.000	nan	0.999	0.987	1.000	0.987	0.000	0.184	0.000	0.991	0.990	0.597	0.978	0.094
AufHi	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.128	0.043	0.000	0.000	0.000	0.000	0.000	0.999	nan	0.777	0.010	0.000	0.524	0.501	0.000	0.874	0.001	0.142	0.002	0.324
Alkoh	0.416	0.340	0.514	0.321	0.620	0.392	0.769	0.604	0.300	0.817	0.000	0.731	0.975	0.465	0.997	0.998	0.987	0.777	nan	0.954	0.124	0.224	0.973	0.653	0.046	0.712	0.940	0.294	0.702
Char1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.068	0.365	0.770	0.544	0.961	0.525	1.000	0.010	0.954	nan	0.000	0.160	0.966	0.436	0.029	0.410	0.581	0.906	0.495
Char2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.002	0.215	0.107	0.987	0.000	0.124	0.000	nan	0.245	0.400	0.044	0.046	0.003	0.920	0.294	0.464
Lich1	0.000	0.000	0.000	0.000	0.169	0.001	0.000	0.000	0.000	0.001	0.287	0.068	0.290	0.339	0.802	0.000	0.000	0.524	0.224	0.160	0.245	nan	0.000	0.000	0.091	0.563	0.000	0.518	0.000
Lich2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.207	0.252	0.000	0.588	0.188	0.957	0.001	0.184	0.501	0.973	0.966	0.400	0.000	nan	0.000	0.820	0.382	0.001	0.853	0.000
Zust1	0.000	0.000	0.000	0.000	0.491	0.191	0.000	0.000	0.000	0.000	0.980	0.000	0.007	0.002	0.606	0.000	0.000	0.000	0.653	0.436	0.044	0.000	0.000	nan	0.001	0.967	0.007	0.207	0.000
Zust2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.390	0.793	0.099	0.940	0.533	0.999	0.076	0.991	0.874	0.046	0.029	0.046	0.091	0.820	0.001	nan	0.592	0.852	0.152	0.002
Fstf	0.502	0.499	0.126	0.342	0.842	0.226	0.840	0.822	0.886	0.000	0.134	0.000	0.001	0.000	0.379	0.988	0.990	0.001	0.712	0.410	0.003	0.563	0.382	0.967	0.592	nan	0.850	0.281	0.819
WoTag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.037	0.631	0.107	0.453	0.344	0.306	0.312	0.597	0.142	0.940	0.581	0.920	0.000	0.001	0.007	0.852	0.850	nan	0.100	0.011
FeiTag	0.930	0.904	0.189	0.125	0.584	0.195	0.800	0.360	0.419	0.738	0.727	0.377	0.932	0.677	0.963	0.055	0.978	0.002	0.294	0.906	0.294	0.518	0.853	0.207	0.152	0.281	0.100	nan	0.092
Month	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.593	0.067	0.055	0.045	0.380	0.000	0.094	0.324	0.702	0.495	0.464	0.000	0.000	0.000	0.002	0.819	0.011	0.092	nan

Table A.44: Significancy matrix for BAYSIS selected data (Jam Follower)

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	NaN	r	r	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TAvg	r	NaN	r	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SMax	r	r	NaN	r	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SAvg	r	r	r	NaN	r	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TDist	r	r	r	r	NaN	r	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
SDist	r	r	r	r	r	NaN	r	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Cov	r	r	r	r	r	r	NaN	r	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLCar	r	r	r	r	r	r	r	NaN	r	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
TLHGV	r	r	r	r	r	r	r	r	NaN	η	η	η	τ	η	η	η	η	η	r_{pq}	η	η	η	η	η	η	τ	η	τ	η
Str	η	η	η	η	η	η	η	η	η	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Kat	η	η	η	η	η	η	η	η	η	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Typ	η	η	η	η	η	η	η	η	η	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
Betei	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt1	η	η	η	η	η	η	η	η	η	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
UArt2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V	V
AUrs2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V	V
AufHi	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V	V
Alkoh	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V	V
Char1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V	V
Char2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V	V
Lich1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V	V
Lich2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V	V
Zust1	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V	V
Zust2	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V	V
Fstf	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V	V
WoTag	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V	V
FeiTag	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN	V
Month	η	η	η	η	η	η	η	η	η	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	NaN

Table A.45: Coefficient matrix for BAYSIS selected data (Jam Follower)

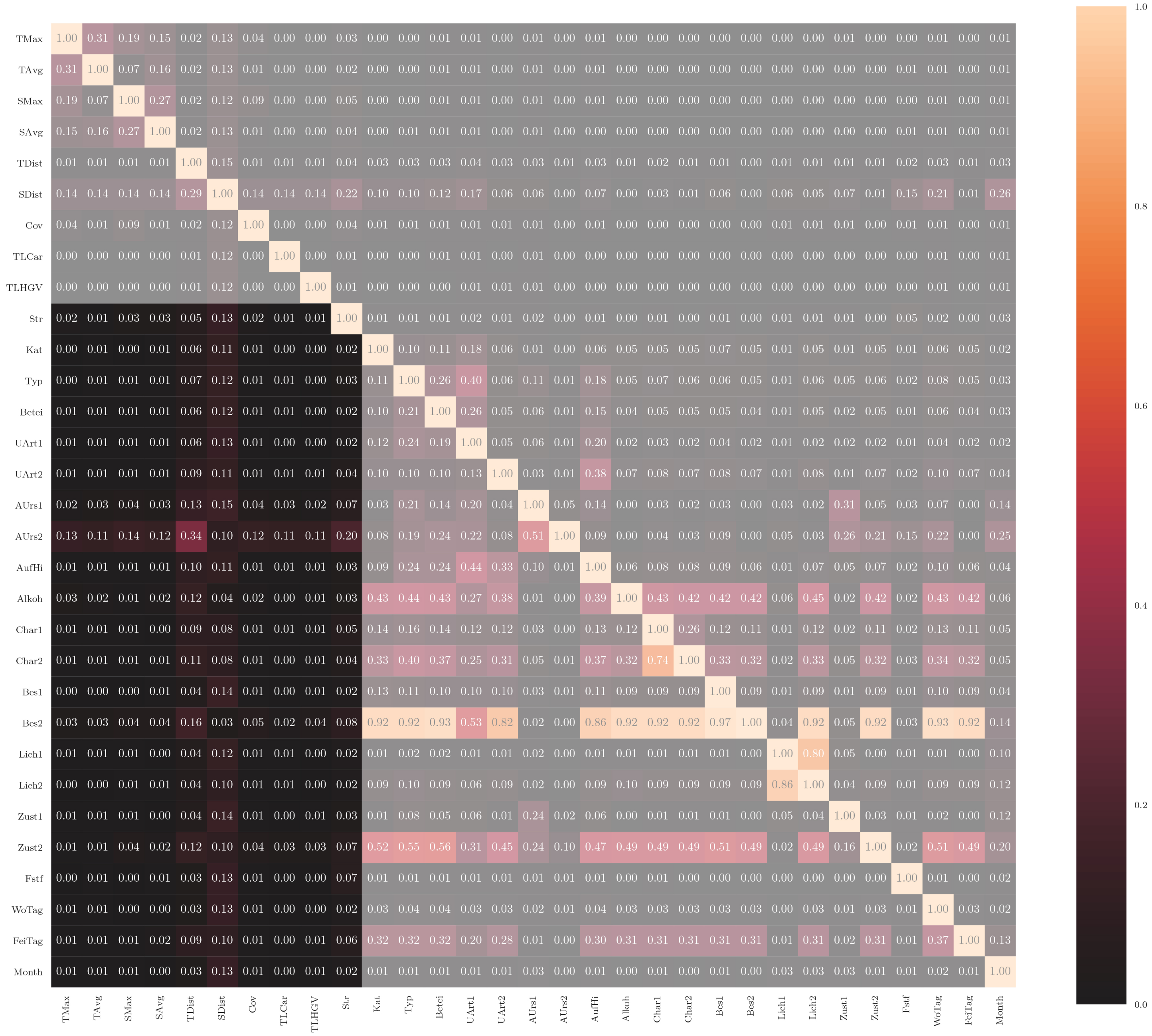


Figure A.4: Predictability matrix for congestion - accident matched data, calculated with Theil's U

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Bes1	Bes2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	1.00	0.31	0.19	0.15	0.02	0.13	0.04	0.00	0.00	0.03	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01
TAvg	0.31	1.00	0.07	0.16	0.02	0.13	0.01	0.00	0.00	0.02	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01
SMax	0.19	0.07	1.00	0.27	0.02	0.12	0.09	0.00	0.00	0.05	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
SAvg	0.15	0.16	0.27	1.00	0.02	0.13	0.01	0.00	0.00	0.04	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01
TDist	0.01	0.01	0.01	0.01	1.00	0.15	0.01	0.01	0.01	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.01	0.03	0.01	0.02	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.03	0.01	0.03
SDist	0.14	0.14	0.14	0.14	0.29	1.00	0.14	0.14	0.14	0.22	0.10	0.10	0.12	0.17	0.06	0.06	0.00	0.07	0.00	0.03	0.01	0.06	0.00	0.06	0.05	0.07	0.01	0.15	0.21	0.01	0.26
Cov	0.04	0.01	0.09	0.01	0.02	0.12	1.00	0.00	0.00	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01
TLCar	0.00	0.00	0.00	0.00	0.01	0.12	0.00	1.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
TLHGV	0.00	0.00	0.00	0.00	0.01	0.12	0.00	0.00	1.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
Str	0.02	0.01	0.03	0.03	0.05	0.13	0.02	0.01	0.01	1.00	0.01	0.01	0.01	0.02	0.01	0.02	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.05	0.02	0.00	0.03
Kat	0.00	0.01	0.00	0.01	0.06	0.11	0.01	0.00	0.00	0.02	1.00	0.10	0.11	0.18	0.06	0.01	0.00	0.06	0.05	0.05	0.05	0.07	0.05	0.01	0.05	0.01	0.05	0.01	0.06	0.05	0.02
Typ	0.00	0.01	0.01	0.01	0.07	0.12	0.01	0.01	0.00	0.03	0.11	1.00	0.26	0.40	0.06	0.11	0.01	0.18	0.05	0.07	0.06	0.06	0.05	0.01	0.06	0.05	0.06	0.02	0.08	0.05	0.03
Betei	0.01	0.01	0.01	0.01	0.06	0.12	0.01	0.01	0.00	0.02	0.10	0.21	1.00	0.26	0.05	0.06	0.01	0.15	0.04	0.05	0.05	0.05	0.04	0.01	0.05	0.02	0.05	0.01	0.06	0.04	0.03
UArt1	0.01	0.01	0.01	0.01	0.06	0.13	0.01	0.00	0.00	0.02	0.12	0.24	0.19	1.00	0.05	0.06	0.01	0.20	0.02	0.03	0.02	0.04	0.02	0.01	0.02	0.02	0.02	0.01	0.04	0.02	0.02
UArt2	0.01	0.01	0.01	0.01	0.09	0.11	0.01	0.01	0.01	0.04	0.10	0.10	0.10	0.13	1.00	0.03	0.01	0.38	0.07	0.08	0.07	0.08	0.07	0.01	0.08	0.01	0.07	0.02	0.10	0.07	0.04
AUrs1	0.02	0.03	0.04	0.03	0.13	0.15	0.04	0.03	0.02	0.07	0.03	0.21	0.14	0.20	0.04	1.00	0.05	0.14	0.00	0.03	0.02	0.03	0.00	0.03	0.02	0.31	0.05	0.03	0.07	0.00	0.14
AUrs2	0.13	0.11	0.14	0.12	0.34	0.10	0.12	0.11	0.11	0.20	0.08	0.19	0.24	0.22	0.08	0.51	1.00	0.09	0.00	0.04	0.03	0.09	0.00	0.05	0.03	0.26	0.21	0.15	0.22	0.00	0.25
AufHi	0.01	0.01	0.01	0.01	0.10	0.11	0.01	0.01	0.01	0.03	0.09	0.24	0.24	0.44	0.33	0.10	0.01	1.00	0.06	0.08	0.08	0.09	0.06	0.01	0.07	0.05	0.07	0.02	0.10	0.06	0.04
Alkoh	0.03	0.02	0.01	0.02	0.12	0.04	0.02	0.00	0.01	0.03	0.43	0.44	0.43	0.27	0.38	0.01	0.00	0.39	1.00	0.43	0.42	0.42	0.42	0.06	0.45	0.02	0.42	0.02	0.43	0.42	0.06
Char1	0.01	0.01	0.01	0.00	0.09	0.08	0.01	0.01	0.01	0.05	0.14	0.16	0.14	0.12	0.12	0.03	0.00	0.13	0.12	1.00	0.26	0.12	0.11	0.01	0.12	0.02	0.11	0.02	0.13	0.11	0.05
Char2	0.01	0.01	0.01	0.01	0.11	0.08	0.01	0.00	0.01	0.04	0.33	0.40	0.37	0.25	0.31	0.05	0.01	0.37	0.32	0.74	1.00	0.33	0.32	0.02	0.33	0.05	0.32	0.03	0.34	0.32	0.05
Bes1	0.00	0.00	0.00	0.01	0.04	0.14	0.01	0.00	0.01	0.02	0.13	0.11	0.10	0.10	0.10	0.03	0.01	0.11	0.09	0.09	0.09	1.00	0.09	0.01	0.09	0.01	0.09	0.01	0.10	0.09	0.04
Bes2	0.03	0.03	0.04	0.04	0.16	0.03	0.05	0.02	0.04	0.08	0.92	0.92	0.93	0.53	0.82	0.02	0.00	0.86	0.92	0.92	0.92	0.97	1.00	0.04	0.92	0.05	0.92	0.03	0.93	0.92	0.14
Lich1	0.01	0.01	0.01	0.00	0.04	0.12	0.01	0.01	0.00	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	1.00	0.80	0.05	0.00	0.01	0.01	0.00	0.10
Lich2	0.00	0.00	0.01	0.01	0.04	0.10	0.01	0.01	0.00	0.02	0.09	0.10	0.09	0.06	0.09	0.02	0.00	0.09	0.10	0.09	0.09	0.09	0.09	0.86	1.00	0.04	0.09	0.01	0.09	0.09	0.12
Zust1	0.01	0.01	0.01	0.00	0.04	0.14	0.01	0.00	0.01	0.03	0.01	0.08	0.05	0.06	0.01	0.24	0.02	0.06	0.00	0.01	0.01	0.01	0.00	0.05	0.04	1.00	0.03	0.01	0.02	0.00	0.12
Zust2	0.01	0.01	0.04	0.02	0.12	0.10	0.04	0.03	0.03	0.07	0.52	0.55	0.56	0.31	0.45	0.24	0.10	0.47	0.49	0.49	0.49	0.51	0.49	0.02	0.49	0.16	1.00	0.02	0.51	0.49	0.20
Fstf	0.00	0.01	0.00	0.01	0.03	0.13	0.01	0.00	0.00	0.07	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.01	0.00	0.02
WoTag	0.01	0.01	0.00	0.00	0.03	0.13	0.01	0.00	0.00	0.02	0.03	0.04	0.04	0.03	0.03	0.02	0.01	0.04	0.03	0.03	0.03	0.03	0.03	0.00	0.03	0.01	0.03	0.01	1.00	0.03	0.02
FeiTag	0.01	0.01	0.01	0.02	0.09	0.10	0.01	0.00	0.01	0.06	0.32	0.32	0.32	0.20	0.28	0.01	0.00	0.30	0.31	0.31	0.31	0.31	0.31	0.01	0.31	0.02	0.31	0.01	0.37	1.00	0.13
Month	0.01	0.01	0.01	0.00	0.03	0.13	0.01	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.03	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.03	0.03	0.03	0.01	0.01	0.02	0.01	1.00

Table A.46: Predictability matrix for congestion - accident matched data, calculated with Theil's U

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	Kat	Typ	Betei	UArt1	UArt2	AUrs1	AUrs2	AufHi	Alkoh	Char1	Char2	Bes1	Bes2	Lich1	Lich2	Zust1	Zust2	Fstf	WoTag	FeiTag	Month
TMax	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
TAvg	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
SMax	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
SAvg	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
TDist	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
SDist	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
Cov	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
TLCar	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
TLHGV	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
Str	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
Kat	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
Typ	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
Betei	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
UArt1	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
UArt2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
AUrs1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
AUrs2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U
AufHi	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U
Alkoh	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U
Char1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U
Char2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U
Bes1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U
Bes2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U
Lich1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U
Lich2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U
Zust1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U
Zust2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U
Fstf	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U
WoTag	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U
FeiTag	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U
Month	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN

Table A.47: Predictability coefficient matrix for congestion - accident matched data

B ArbIS Figures, Tables and Listings

	Strasse	AnzGesperrtFs	Einzug	Richtung	Length	Duration	Month
Strasse	1.00	0.16	0.17	0.05	0.04	0.02	0.06
AnzGesperrtFs	0.16	1.00	0.50	0.00	0.02	0.11	0.06
Einzug	0.17	0.50	1.00	0.02	-0.00	-0.15	0.11
Richtung	0.05	0.00	0.02	1.00	-0.02	-0.00	0.03
Length	0.04	0.02	-0.00	-0.02	1.00	0.08	0.04
Duration	0.02	0.11	-0.15	-0.00	0.08	1.00	0.02
Month	0.06	0.06	0.11	0.03	0.04	0.02	1.00

Table B.1: Correlation matrix for ArbIS dataset, calculated with Cramer's V, η, τ, r_{pq}, r

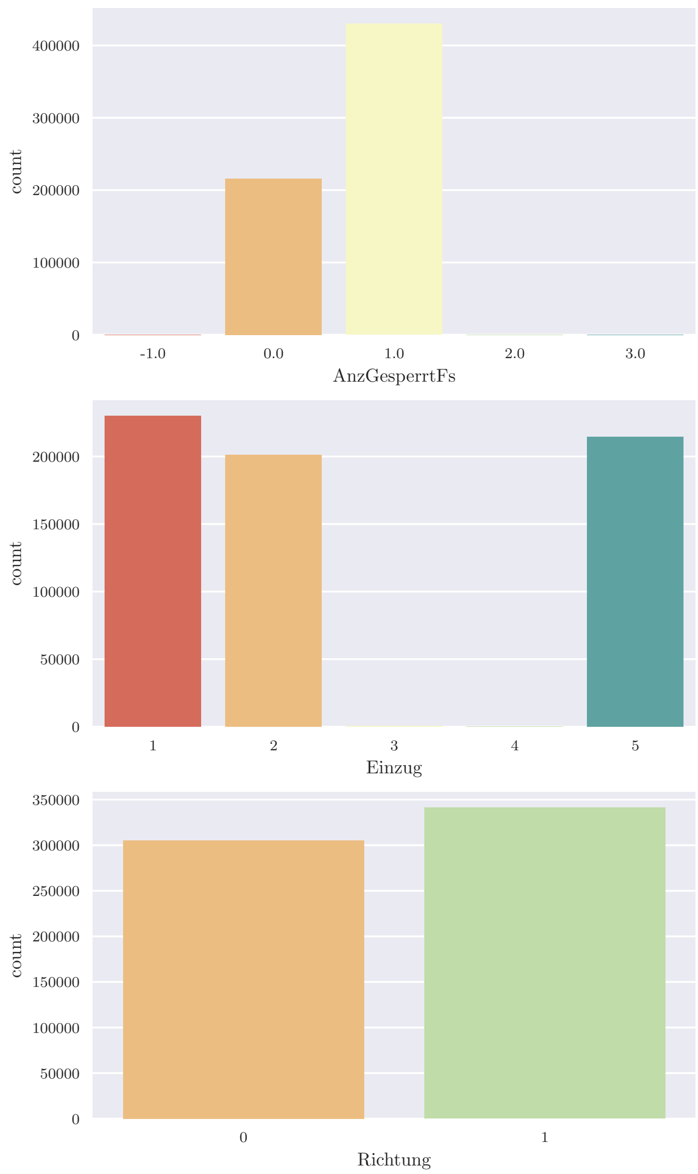


Figure B.1: Distribution of the accident category *AnzGesperrtFs*, *Einzug* and *Richtung*

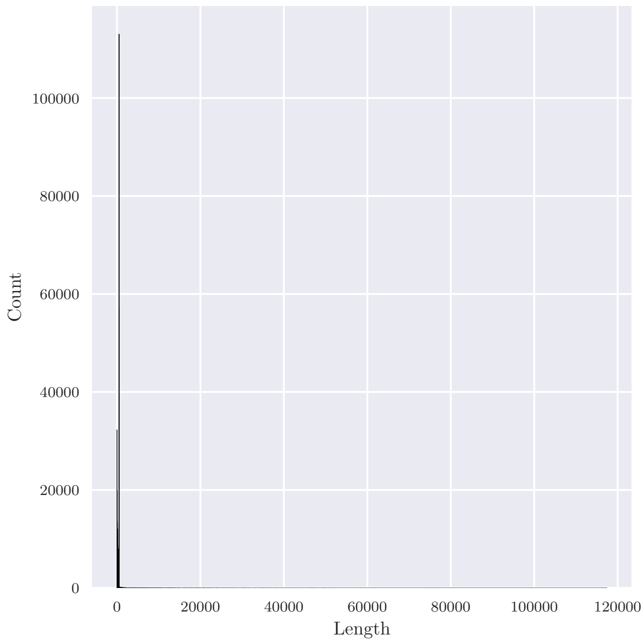


Figure B.2: Distribution of the roadwork parameter *Length*

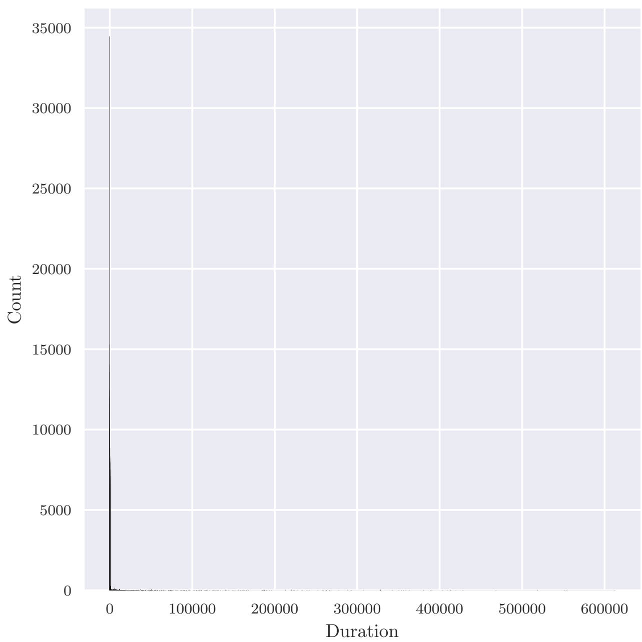


Figure B.3: Distribution of the roadwork parameter *Duration*

	Strasse	AnzGesperrtFs	Einzug	Richtung	Length	Duration	Month
Strasse	NaN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0
AnzGesperrtFs	0.0	NaN	0.0000	0.2547	0.0000	0.0000	0.0
Einzug	0.0	0.0000	NaN	0.0000	0.0006	0.0000	0.0
Richtung	0.0	0.2547	0.0000	NaN	0.0000	0.0489	0.0
Length	0.0	0.0000	0.0006	0.0000	NaN	0.0000	0.0
Duration	0.0	0.0000	0.0000	0.0489	0.0000	NaN	0.0
Month	0.0	0.0000	0.0000	0.0000	0.0000	0.0000	NaN

Table B.2: Significancy matrix for ArbIS dataset

	Strasse	AnzGesperrtFs	Einzug	Richtung	Length	Duration	Month
Strasse	NaN	V	V	V	η	η	V
AnzGesperrtFs	V	NaN	V	V	τ	τ	V
Einzug	V	V	NaN	V	τ	τ	V
Richtung	V	V	V	NaN	r_{pq}	r_{pq}	V
Length	η	τ	τ	r_{pq}	NaN	r	η
Duration	η	τ	τ	r_{pq}	r	NaN	η
Month	V	V	V	V	η	η	NaN

Table B.3: Coefficient matrix for ArbIS dataset

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	1.00															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	1.00	1.00	1.00	1.00	1.00											
A3	0.00	0.04	1.00	1.00	0.01	1.00										
A99	1.00	1.00	1.00	1.00	0.48	1.00	1.00									
A96	1.00	0.82	1.00	1.00	1.00	1.00	0.00	0.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	0.79	0.74	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.00	1.00	0.09	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	0.68	1.00	1.00	1.00	0.01	0.48	1.00	1.00	1.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.4: Pairwise Wilcoxon T-test for *Strasse* and *Maximal Temporal Extent* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	1.00															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	1.00	1.00	1.00	1.00	1.00											
A3	1.00	1.00	1.00	1.00	0.69	1.00										
A99	0.00	0.00	0.08	1.00	1.00	1.00	0.00									
A96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.55								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.04	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.00	1.00	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	0.14	1.00	1.00	0.00	0.63	0.53	1.00	1.00	0.02	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.5: Pairwise Wilcoxon T-test for *Strasse* and *Average Temporal Extent* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	0.58															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	1.00	1.00	1.00	1.00	1.00											
A3	0.00	0.00	0.00	0.52	0.01	0.05										
A99	0.07	0.00	0.01	0.21	0.26	0.20	1.00									
A96	0.00	0.06	1.00	1.00	0.00	1.00	0.00	0.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	0.11	0.03	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.04	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.81	1.00					
A93	0.00	0.00	0.02	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.73				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	1.00	1.00	0.04	0.04	0.89	1.00	1.00	1.00	0.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.6: Pairwise Wilcoxon T-test for *Strasse* and *Maximal Spatial Extent* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	0.06															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	0.62	1.00	1.00	1.00	1.00											
A3	1.00	0.00	1.00	1.00	1.00	0.02										
A99	0.01	1.00	1.00	1.00	0.75	1.00	0.00									
A96	0.00	1.00	1.00	1.00	0.44	1.00	0.00	1.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A92	1.00	0.62	1.00	1.00	1.00	0.73	1.00	0.10	0.14	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.13	1.00	1.00	0.00	1.00	0.00	0.11	1.00	1.00	0.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.7: Pairwise Wilcoxon T-test for *Strasse* and *Average Spatial Extent* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	0.00															
A70	0.57	1.00														
A71	1.00	1.00	1.00													
A6	0.10	1.00	1.00	1.00												
A73	1.00	1.00	1.00	1.00	1.00											
A3	0.00	1.00	1.00	1.00	1.00	0.85										
A99	0.38	1.00	1.00	1.00	1.00	1.00	0.26									
A96	1.00	1.00	1.00	1.00	1.00	1.00	0.03	1.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A92	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.63	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00			
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00		1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00		1.00	

Table B.8: Pairwise Wilcoxon T-test for *Strasse* and *Temporal Distance* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	0.62															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	0.03	1.00	1.00	1.00	0.89											
A3	1.00	1.00	1.00	1.00	1.00	0.33										
A99	1.00	0.61	1.00	1.00	1.00	0.03	1.00									
A96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.05	1.00	1.00	0.01	1.00	0.00	0.00	0.06	0.16	0.12	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00			
A94	1.00	0.09	0.72	1.00	0.76	0.00	0.40	1.00	0.30	1.00	1.00	1.00	0.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		1.00		1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.9: Pairwise Wilcoxon T-test for *Strasse* and *Spatial Distance* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	0.07															
A70	0.01	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	1.00	1.00	1.00	1.00	1.00											
A3	0.00	0.00	0.00	1.00	0.01	1.00										
A99	0.00	0.00	0.00	0.15	0.00	0.00	0.00									
A96	0.00	0.18	1.00	1.00	0.00	0.00	0.00	0.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.02	1.00							
A92	0.00	1.00	1.00	1.00	0.04	0.05	0.00	0.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.36	1.00	1.00	1.00	0.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.10: Pairwise Wilcoxon T-test for *Strasse* and *Coverage* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	1.00															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	0.56	0.08	1.00	1.00												
A73	1.00	1.00	1.00	1.00	1.00											
A3	1.00	1.00	1.00	1.00	1.00	1.00										
A99	1.00	1.00	1.00	1.00	1.00	1.00	1.00									
A96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.00	1.00	1.00	0.10	0.52	0.00	0.00	0.00	1.00	0.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.11: Pairwise Wilcoxon T-test for *Strasse* and *Time-loss Car* complete

	A9	A7	A70	A71	A6	A73	A3	A99	A96	A995	A92	A72	A93	A95	A94	A980
A7	1.00															
A70	1.00	1.00														
A71	1.00	1.00	1.00													
A6	1.00	1.00	1.00	1.00												
A73	0.22	1.00	1.00	1.00	1.00											
A3	1.00	1.00	1.00	1.00	1.00	1.00										
A99	0.10	1.00	1.00	1.00	1.00	1.00	1.00									
A96	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00								
A995	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
A92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
A72	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
A93	0.00	0.00	0.00	0.62	0.00	0.01	0.00	0.00	0.00	1.00	0.00	1.00				
A95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
A94	1.00	1.00	1.00	1.00	1.00	0.57	1.00	0.79	1.00	1.00	1.00	1.00	0.00	1.00		
A980	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
A45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.12: Pairwise Wilcoxon T-test for *Strasse* and *Time-loss HGV* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	1.00	1.00									
Apr	1.00	1.00	1.00								
May	1.00	0.17	1.00	0.26							
Jun	1.00	1.00	1.00	1.00	1.00						
Jul	1.00	0.68	1.00	1.00	1.00	1.00					
Aug	1.00	1.00	1.00	1.00	0.89	1.00	1.00				
Sep	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
Oct	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00		
Nov	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	
Dec	1.00	1.00	0.97	1.00	0.01	1.00	0.20	1.00	0.00	1.00	1.00

Table B.13: Pairwise Wilcoxon T-test for *Month* and *Average Temporal Extent* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	0.41	0.60									
Apr	1.00	1.00	0.01								
May	1.00	1.00	1.00	1.00							
Jun	1.00	1.00	1.00	0.30	1.00						
Jul	1.00	1.00	0.37	1.00	1.00	1.00					
Aug	0.00	0.00	0.69	0.00	0.00	1.00	0.00				
Sep	0.37	0.01	1.00	0.00	0.07	1.00	0.00	1.00			
Oct	0.02	0.00	1.00	0.00	0.01	1.00	0.00	1.00	1.00		
Nov	0.00	0.00	0.02	0.00	0.00	0.01	0.00	1.00	0.14	1.00	
Dec	0.68	0.30	1.00	0.01	1.00	1.00	0.01	1.00	1.00	1.00	0.05

Table B.14: Pairwise Wilcoxon T-test for *Month* and *Maximal Spatial Extent* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	1.00	0.49									
Apr	1.00	1.00	1.00								
May	1.00	0.82	1.00	1.00							
Jun	1.00	0.47	1.00	1.00	1.00						
Jul	1.00	0.15	1.00	0.94	1.00	1.00					
Aug	0.25	0.00	1.00	0.00	1.00	1.00	0.51				
Sep	1.00	0.12	1.00	0.29	1.00	1.00	1.00	1.00			
Oct	1.00	0.03	1.00	0.03	1.00	1.00	1.00	1.00	1.00		
Nov	0.00	0.00	0.02	0.00	0.01	0.15	0.00	1.00	0.01	0.09	
Dec	1.00	0.30	1.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00	0.09

Table B.15: Pairwise Wilcoxon T-test for *Month* and *Average Spatial Extent* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	1.00	0.27									
Apr	1.00	0.25	1.00								
May	1.00	0.10	1.00	1.00							
Jun	1.00	1.00	1.00	1.00	1.00						
Jul	1.00	1.00	0.18	0.12	0.02	1.00					
Aug	0.36	0.01	1.00	1.00	1.00	1.00	0.00				
Sep	0.21	0.00	1.00	1.00	1.00	1.00	0.00	1.00			
Oct	1.00	0.05	1.00	1.00	1.00	1.00	0.00	1.00	1.00		
Nov	0.00	0.00	0.03	0.01	0.03	0.01	0.00	0.68	0.73	0.09	
Dec	0.01	0.00	0.20	0.10	0.24	0.04	0.00	1.00	1.00	0.68	1.00

Table B.16: Pairwise Wilcoxon T-test for *Month* and *Spatial Distance* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	1.00	1.00									
Apr	1.00	1.00	0.75								
May	1.00	1.00	0.71	1.00							
Jun	1.00	1.00	1.00	1.00	1.00						
Jul	1.00	0.79	0.09	1.00	1.00	0.61					
Aug	0.42	0.32	0.11	0.00	0.00	0.35	0.00				
Sep	0.14	0.03	0.03	0.00	0.00	0.02	0.00	1.00			
Oct	0.50	0.35	0.20	0.00	0.00	0.14	0.00	1.00	1.00		
Nov	0.06	0.01	0.00	0.00	0.00	0.03	0.00	0.35	1.00	1.00	
Dec	1.00	1.00	1.00	0.01	0.00	1.00	0.00	1.00	1.00	1.00	0.79

Table B.17: Pairwise Wilcoxon T-test for *Month* and *Coverage* complete

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Feb	1.00										
Mar	0.05	1.00									
Apr	0.01	1.00	1.00								
May	0.85	1.00	1.00	1.00							
Jun	1.00	1.00	1.00	0.46	1.00						
Jul	0.05	1.00	1.00	1.00	1.00	1.00					
Aug	0.22	1.00	1.00	1.00	1.00	1.00	1.00				
Sep	1.00	1.00	0.08	0.00	1.00	1.00	0.00	0.04			
Oct	0.04	1.00	1.00	1.00	1.00	0.40	0.73	1.00	0.00		
Nov	1.00	1.00	0.10	0.03	1.00	1.00	0.00	0.16	1.00	0.00	
Dec	1.00	1.00	0.08	0.00	1.00	1.00	0.04	0.28	1.00	0.02	1.00

Table B.18: Pairwise Wilcoxon T-test for *Month* and *Time-loss Car* complete

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	AGF	Einzug	Richtung	Length	Duration	Month
TMax	1.00	0.80	0.54	0.47	-0.18	-0.02	-0.18	0.06	0.00	0.20	-0.04	0.02	0.02	0.06	0.02	0.13
TAvg	0.80	1.00	0.24	0.43	-0.17	-0.00	0.20	0.04	0.03	0.19	0.05	-0.07	0.02	-0.00	0.02	0.20
SMax	0.54	0.24	1.00	0.72	-0.14	-0.08	-0.48	0.00	-0.04	0.29	-0.11	0.12	-0.01	0.13	0.00	0.20
SAvg	0.47	0.43	0.72	1.00	-0.16	-0.11	0.01	0.00	-0.00	0.20	-0.04	0.04	-0.02	0.07	0.00	0.14
TDist	-0.18	-0.17	-0.14	-0.16	1.00	0.07	-0.02	0.00	0.02	0.16	-0.01	0.00	0.01	-0.06	-0.02	0.15
SDist	-0.02	-0.00	-0.08	-0.11	0.07	1.00	-0.07	0.01	0.01	0.16	-0.07	0.09	0.03	-0.11	-0.01	0.13
Cov	-0.18	0.20	-0.48	0.01	-0.02	-0.07	1.00	-0.07	-0.03	0.40	0.19	-0.16	-0.00	-0.11	-0.01	0.22
TLCar	0.06	0.04	0.00	0.00	0.00	0.01	-0.07	1.00	0.10	0.14	-0.04	0.01	-0.02	0.02	0.00	0.14
TLHGV	0.00	0.03	-0.04	-0.00	0.02	0.01	-0.03	0.10	1.00	0.16	-0.02	0.01	0.03	0.00	0.02	0.12
Str	0.20	0.19	0.29	0.20	0.16	0.16	0.40	0.14	0.16	1.00	0.18	0.17	0.13	0.17	0.07	0.18
AGF	-0.04	0.05	-0.11	-0.04	-0.01	-0.07	0.19	-0.04	-0.02	0.18	1.00	-0.73	0.06	-0.05	-0.07	0.13
Einzug	0.02	-0.07	0.12	0.04	0.00	0.09	-0.16	0.01	0.01	0.17	-0.73	1.00	0.14	0.03	-0.13	0.14
Richtung	0.02	0.02	-0.01	-0.02	0.01	0.03	-0.00	-0.02	0.03	0.13	0.06	0.14	1.00	-0.05	-0.07	0.14
Length	0.06	-0.00	0.13	0.07	-0.06	-0.11	-0.11	0.02	0.00	0.17	-0.05	0.03	-0.05	1.00	0.07	0.09
Duration	0.02	0.02	0.00	0.00	-0.02	-0.01	-0.01	0.00	0.02	0.07	-0.07	-0.13	-0.07	0.07	1.00	0.05
Month	0.13	0.20	0.20	0.14	0.15	0.13	0.22	0.14	0.12	0.18	0.13	0.14	0.14	0.09	0.05	1.00

Table B.19: Correlation matrix for ArbIS matched data, calculated with Cramer's V, η, τ, r_{pq}, r

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	AGF	Einzug	Richtung	Length	Duration	Month
TMax	nan	0.000	0.000	0.000	0.000	0.385	0.000	0.001	0.965	0.000	0.024	0.107	0.267	0.001	0.173	0.000
TAvg	0.000	nan	0.000	0.000	0.000	0.953	0.000	0.047	0.071	0.000	0.011	0.000	0.301	0.987	0.340	0.000
SMax	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.859	0.034	0.000	0.000	0.000	0.451	0.000	0.993	0.000
SAvg	0.000	0.000	0.000	nan	0.000	0.000	0.663	0.988	0.896	0.000	0.041	0.004	0.397	0.000	0.934	0.000
TDist	0.000	0.000	0.000	0.000	nan	0.000	0.201	0.819	0.190	0.000	0.452	0.765	0.506	0.001	0.163	0.000
SDist	0.385	0.953	0.000	0.000	0.000	nan	0.000	0.527	0.445	0.000	0.000	0.000	0.063	0.000	0.433	0.000
Cov	0.000	0.000	0.000	0.663	0.201	0.000	nan	0.000	0.050	0.000	0.000	0.000	0.799	0.000	0.508	0.000
TLCar	0.001	0.047	0.859	0.988	0.819	0.527	0.000	nan	0.000	0.000	0.032	0.350	0.229	0.183	0.913	0.000
TLHGV	0.965	0.071	0.034	0.896	0.190	0.445	0.050	0.000	nan	0.000	0.237	0.677	0.112	0.802	0.279	0.000
Str	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan	0.000	0.000	0.000	0.000	0.000	0.000
AGF	0.024	0.011	0.000	0.041	0.452	0.000	0.000	0.032	0.237	0.000	nan	0.000	0.000	0.002	0.000	0.000
Einzug	0.107	0.000	0.000	0.004	0.765	0.000	0.000	0.350	0.677	0.000	0.000	nan	0.000	0.046	0.000	0.000
Richtung	0.267	0.301	0.451	0.397	0.506	0.063	0.799	0.229	0.112	0.000	0.000	0.000	nan	0.008	0.000	0.000
Length	0.001	0.987	0.000	0.000	0.001	0.000	0.000	0.183	0.802	0.000	0.002	0.046	0.008	nan	0.000	0.000
Duration	0.173	0.340	0.993	0.934	0.163	0.433	0.508	0.913	0.279	0.000	0.000	0.000	0.000	0.000	nan	0.000
Month	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	nan

Table B.20: Significancy matrix for ArbIS matched data

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	AGF	Einzug	Richtung	Length	Duration	Month
TMax	NaN	r	r	r	r	r	r	r	r	η	r	τ	r_{pq}	r	r	η
TAvg	r	NaN	r	r	r	r	r	r	r	η	r	τ	r_{pq}	r	r	η
SMax	r	r	NaN	r	r	r	r	r	r	η	r	τ	r_{pq}	r	r	η
SAvg	r	r	r	NaN	r	r	r	r	r	η	r	τ	r_{pq}	r	r	η
TDist	r	r	r	r	NaN	r	r	r	r	η	r	τ	r_{pq}	r	r	η
SDist	r	r	r	r	r	NaN	r	r	r	η	r	τ	r_{pq}	r	r	η
Cov	r	r	r	r	r	r	NaN	r	r	η	r	τ	r_{pq}	r	r	η
TLCar	r	r	r	r	r	r	r	NaN	r	η	r	τ	r_{pq}	r	r	η
TLHGV	r	r	r	r	r	r	r	r	NaN	η	r	τ	r_{pq}	r	r	η
Str	η	η	η	η	η	η	η	η	η	NaN	η	V	V	η	η	V
AGF	r	r	r	r	r	r	r	r	r	η	NaN	τ	r_{pq}	r	r	η
Einzug	τ	τ	τ	τ	τ	τ	τ	τ	τ	V	τ	NaN	V	τ	τ	V
Richtung	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	r_{pq}	V	r_{pq}	V	NaN	r_{pq}	r_{pq}	V
Length	r	r	r	r	r	r	r	r	r	η	r	τ	r_{pq}	NaN	r	η
Duration	r	r	r	r	r	r	r	r	r	η	r	τ	r_{pq}	r	NaN	η
Month	η	η	η	η	η	η	η	η	η	V	η	V	V	η	η	NaN

Table B.21: Coefficient matrix for ArbIS matched data

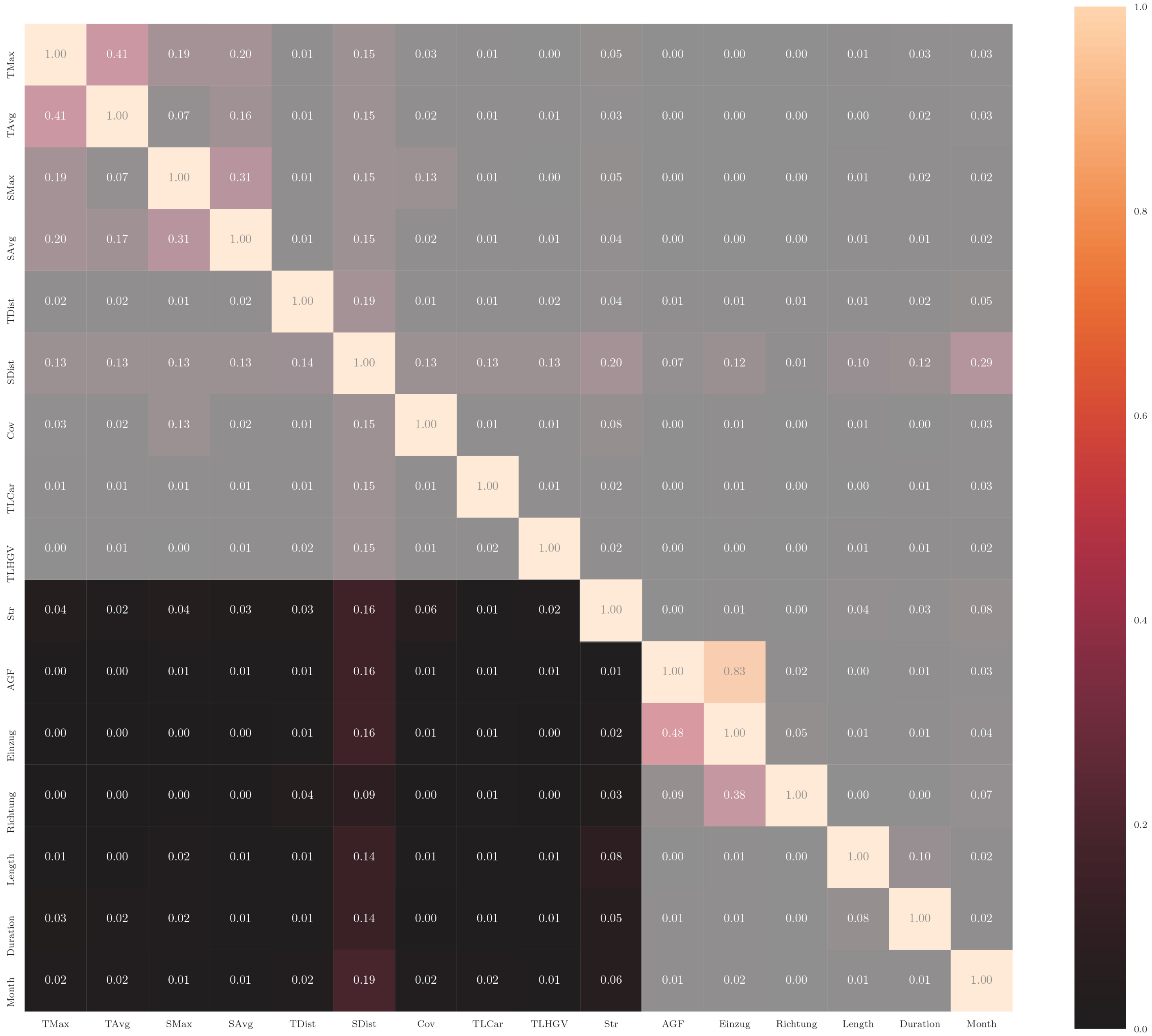


Figure B.4: Predictability matrix for congestion - roadwork matched data, calculated with Theil's U

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	AGF	Einzug	Richtung	Length	Duration	Month
TMax	1.00	0.41	0.19	0.20	0.01	0.15	0.03	0.01	0.00	0.05	0.00	0.00	0.00	0.01	0.03	0.03
TAvg	0.41	1.00	0.07	0.16	0.01	0.15	0.02	0.01	0.01	0.03	0.00	0.00	0.00	0.00	0.02	0.03
SMax	0.19	0.07	1.00	0.31	0.01	0.15	0.13	0.01	0.00	0.05	0.00	0.00	0.00	0.01	0.02	0.02
SAvg	0.20	0.17	0.31	1.00	0.01	0.15	0.02	0.01	0.01	0.04	0.00	0.00	0.00	0.01	0.01	0.02
TDist	0.02	0.02	0.01	0.02	1.00	0.19	0.01	0.01	0.02	0.04	0.01	0.01	0.01	0.01	0.02	0.05
SDist	0.13	0.13	0.13	0.13	0.14	1.00	0.13	0.13	0.13	0.20	0.07	0.12	0.01	0.10	0.12	0.29
Cov	0.03	0.02	0.13	0.02	0.01	0.15	1.00	0.01	0.01	0.08	0.00	0.01	0.00	0.01	0.00	0.03
TLCar	0.01	0.01	0.01	0.01	0.01	0.15	0.01	1.00	0.01	0.02	0.00	0.01	0.00	0.00	0.01	0.03
TLHGV	0.00	0.01	0.00	0.01	0.02	0.15	0.01	0.02	1.00	0.02	0.00	0.00	0.00	0.01	0.01	0.02
Str	0.04	0.02	0.04	0.03	0.03	0.16	0.06	0.01	0.02	1.00	0.00	0.01	0.00	0.04	0.03	0.08
AGF	0.00	0.00	0.01	0.01	0.01	0.16	0.01	0.01	0.01	0.01	1.00	0.83	0.02	0.00	0.01	0.03
Einzug	0.00	0.00	0.00	0.00	0.01	0.16	0.01	0.01	0.00	0.02	0.48	1.00	0.05	0.01	0.01	0.04
Richtung	0.00	0.00	0.00	0.00	0.04	0.09	0.00	0.01	0.00	0.03	0.09	0.38	1.00	0.00	0.00	0.07
Length	0.01	0.00	0.02	0.01	0.01	0.14	0.01	0.01	0.01	0.08	0.00	0.01	0.00	1.00	0.10	0.02
Duration	0.03	0.02	0.02	0.01	0.01	0.14	0.00	0.01	0.01	0.05	0.01	0.01	0.00	0.08	1.00	0.02
Month	0.02	0.02	0.01	0.01	0.02	0.19	0.02	0.02	0.01	0.06	0.01	0.02	0.00	0.01	0.01	1.00

Table B.22: Predictability matrix for congestion - roadwork matched data, calculated with Theil's U

	TMax	TAvg	SMax	SAvg	TDist	SDist	Cov	TLCar	TLHGV	Str	AGF	Einzug	Richtung	Length	Duration	Month
TMax	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
TAvg	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U	U
SMax	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U	U
SAvg	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U	U
TDist	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U	U
SDist	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U	U
Cov	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U	U
TLCar	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U	U
TLHGV	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U	U
Str	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U	U
AGF	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U	U
Einzug	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U	U
Richtung	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U	U
Length	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U	U
Duration	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN	U
Month	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	NaN

Table B.23: Predictability coefficient matrix for congestion - roadwork matched data