

Faktoren des Fahrrad Verkehrs in Mannheim

Maximilian Samuel Weinhold

Economics, 6. Semester

505314

mweinhol@uni-muenster.de

Hausarbeit im Rahmen des Seminars zur

Analyse von Fahrrad-Verkehrsdaten

Sommersemester 2021

Institut für Verkehrswissenschaft

Dr. Jan Wessel

INHALTSVERZEICHNIS

1. <i>Einleitung</i>	4
2. <i>Literaturüberblick</i>	5
3. <i>Beschreibung der Daten</i>	6
4. <i>Deskriptive Analyse</i>	12
4.1 Regressionsanalyse	12
4.2 Negativ Binomial Regression	15
4.3 Utilitaristischer und Freizeitverkehr	16
5. <i>Eigene Forschungsfragen</i>	20
5.1 Model Selektion	20
5.2 Fahrrad Topographie	26
5.3 Vorhersagevergleich	29
6. <i>Fazit</i>	31
7. <i>Anhang</i>	33
<i>Literaturverzeichnis</i>	39

ABBILDUNGSVERZEICHNIS

3.1 Ausprägungen des Fahrradaufkommens	9
--	---

3.2	Zählstationen im Überblick	10
3.3	Boxplots zu den Zählstationen	11
4.1	Fahrräder in Entfernung zur Universität Mannheim Graphisches Beispiel einer Regression	13
5.1	Model Selektion	22
5.2	Fahrrad Topographie Mannheims an einem vergangenen Tag .	29
5.3	Vergleich zur Hochrechnung an der Renzstraße 2021	30
7.1	Korrelations Plot	35
7.2	Monatliche Durchschnittliche Temperaturen und maximale Nie- derschläge nach Tageszeiten 2018	36
7.3	Tägliche Durchschnitts Temperaturen und monatliche maxi- male Niederschläge nach Tageszeiten 2018	37
7.4	Entscheidungsbaum	38

TABELLENVERZEICHNIS

3.1	Datensatz Übersicht	8
4.1	Log-Lineares Regressionsmodell	14
4.2	Negativ Binomial Regression	16
4.3	Utilitaristischer und Freizeitverkehr der Zählstationen	18
4.4	Freizeit- und Berufsverkehr im Vergleich	19
5.1	Modell Vergleich	23
5.2	Verteilung der Vorhersagen	27
5.3	Finales Modell: Model 23	28
7.1	Wetterdaten als täglichen Niederschlag und Höchsttemperatur	33
7.2	Beispielhafte Vorhersagen	34

1. EINLEITUNG

Die mobile Infrastruktur urbaner Zentren in Deutschland unterliegt einem Wandel, der nicht mehr allein das Ziel einer autofreundlicheren Stadt verfolgt, sondern auch andere Verkehrsteilnehmer hervor hebt. Im Besonderen betrifft diese Entwicklung das Fahrrad. So ist laut Nobis (2019) die Anzahl der Fahrradfahrer innerhalb von 15 Jahren gestiegen sowie die Länge der von ihnen zurückgelegten Wege. Ein Beispiel für diese Entwicklungen ist Mannheim. Denn hier verfolgt die städtische Planung ein 21 Punkte Programm (Mannheim (2010)), dass dazu dienen soll den Anteil des Radverkehrs zu steigern.

Im Rahmen dieses Programms installierte die Stadt an verschiedenen Stellen acht Fahrradzähler, deren stündliche Zählungen öffentlich abrufbar sind. Mithilfe dieser Daten lässt sich ein Regressionsmodell entwickeln, dass die kausalen Einflüsse auf den Radverkehr erklärt und das Aufkommen von Fahrrädern vorhersagen kann.

Nach einem Überblick über die aktuelle Literatur zu diesem Thema und einer genaueren Beschreibung der verwendeten Daten möchte diese Hausarbeit in zwei Abschnitten ein statistisches Modell entwickeln. Im ersten Abschnitt liegt der Fokus auf die kausalen Zusammenhänge. Im zweiten Teil geht es um die eigene Forschungsfrage: welches Modell kann man wählen, um die Vorhersage Kraft zu maximieren? Im besten Fall ließe sich das Modell extrapolieren, sodass eine Topographie des Fahrradverkehrs absehbar wäre auch an Stellen ohne Zählstationen.

Diese Hausarbeit baut vornehmlich auf Wessel (2020) und seine Methodik auf, der primär untersuchte welchen Einfluss Wettervorhersagen und tatsächliches Wetter auf den Radverkehr in verschiedenen deutschen Städten haben.

2. LITERATURÜBERBLICK

Da diese Arbeit im Rahmen eines Seminares bei Dr. Jan Wessel entstanden ist, verfolgt sie eine ähnliche Methodik wie bei Wessel (2020). Dort wird der Frage nachgegangen, wie sehr Wettervorhersagen und tatsächliches Wetter einen Einfluss auf das Aufkommen der Fahrradfahrer hat. Dazu entwickelt er ein Log-lineares Regressionsmodell, das nicht nur die notwendigen Wetterdaten berücksichtigt, sondern auch andere wesentliche Effekte wie Feiertage, Schul- und Semesterferien. Sein Modell kommt zu einem adjusted R^2 von 78 %. Die Daten für das Modell stammen von 188 Zählstationen in ganz Deutschland. Daneben untersuchen auch andere Forscher Fragen zum Thema des Verkehrsaufkommens von Fahrrädern. Z.b. untersuchen Zhao et al. (2018) und Hong (2022) den Einfluss von Feinstaubbelastungen auf den Fahrradverkehr. Dabei verwenden diese, ähnlich wie Wessel (2020), Zeitreihendaten. Aber nicht nur Einflüsse äußerer und zeitlicher Umstände wie Wetter oder Feinstaub sind interessante Variablen. Hankey et al. (2021) gehen im Speziellen auf räumliche Daten ein. Sie beobachten über 4000 Zählstellen in den USA und verwenden Google Streetview Bilder und einen Pixelerkennungsalgorithmus. Ein anderes interessantes Forschungsthema, das in Verbindung zu Fahrraddaten steht, ist die Vorhersage von Verkehrsunfällen, in die Fahrradfahrer verwickelt sind. Diesem Thema gehen Prati et al. (2017) nach. Sie greifen hierbei nicht auf Regressionsmodelle zurück, sondern auf CHAID (Chi-square Automatic Interaction Detectors) Entscheidungsbäume und auf Bayesian Netzwerkanalysen. Entscheidungsbäume sind für dieses Thema ebenfalls eine interessante Methode. Im Anhang in der Darstellung 7.4 ist ein Entscheidungsbaum, beruhend auf den Datensatz aus Mannheim, dargestellt. Dennoch verwenden wir folgend ausschließlich Regressionsmodelle.

3. BESCHREIBUNG DER DATEN

Die Bearbeitung des Modells baut überwiegend auf zwei Datenquellen auf. Am wichtigsten sind die Fahrradzählstationen Mannheims, die öffentlich im Netz zugänglich sind (Link: <https://mannheim.opendatasoft.com/page/home/>). Hier sind stündliche Werte gezählter Fahrräder verfügbar an acht verschiedenen Stellen in der Stadt sowie Längen und Breitengrad der Position, Standortname, Zeit und Datum. Im zweiten Schritt werden diese Daten mit stündlichen Daten des Deutschen Wetter Dienstes zur Lufttemperatur in 2 Meter Höhe, zur relativen Feuchte, zum Bedeckungsgrad, zum Niederschlag und zur Sonnenstrahlung verbunden.

Zhao et al. (2018) und Hong (2022) zeigen, dass auch die Feinstaubbelastung den Radverkehr beeinflussen kann. Grundsätzlich wären die dazu notwendigen Daten aus Mannheim über [Openseamap.org](https://openseamap.org) zugänglich, jedoch nicht für den gesamten Betrachtungszeitraum. Da Zhao et al. (2018) und Hong (2022) ihre Forschung in ostasiatischen Ballungsräumen betrieben haben und in einer Stadt wie Mannheim nicht die selbe Feinstaubbelastung zu erwarten ist, kann man von dieser Variable nicht erhoffen, dass sie die notwendige Erklärungskraft bringt, um einen Wegfall des Betrachtungszeitraumes zu rechtfertigen, weshalb Feinstaubmessungen keinen Weg in den Datensatz gefunden haben.

Dafür ist es wichtig Feiertage, Schulferien und Semesterferien zu berücksichtigen. Hier könnte potentiell eine geographische Besonderheit Mannheims eine Rolle spielen, denn das Stadtgebiet grenzt direkt zu zwei verschiedenen Bundesländern an. Zum einem liegt auf dem gegenüberliegenden Rheinufer die Stadt Ludwigshafen, die bereits in Rheinland Pfalz liegt. Zwischen beiden Städten herrscht ein reger Verkehr, weshalb anzunehmen ist, dass voneinander abweichende Feiertage in Rheinland Pfalz und Baden Württemberg eine

Rolle spielen könnten. Ebenso grenzt Hessen an Mannheim an. Jedoch nicht in einer Reichweite, die für Fahrradfahrer realistisch erscheint und nicht mit einer Stadt, die vergleichbar groß wie Ludwigshafen wäre. Somit finden Feiertage für Hessen keinen Weg ins Model.

Dies gilt ebenso für die Schulferien. Es ist anzunehmen, dass innerhalb Mannheims, Schulen überwiegend von Mannheimern besucht werden, weshalb nur die Schulferien von Baden Württemberg im Datensatz zu finden sind. Neben Schülern fahren ebenso häufig Studenten Fahrrad. Die größte Hochschulbildungseinrichtung ist die Universität Mannheim, deren Semesterferien eine signifikante Rolle spielen dürften.

Fügt man all diese Daten zusammen, erhält man einen Datensatz der 410891 Beobachtungen umfasst. Die Ausprägungen pro Tag sind der Darstellung 3.1 zu entnehmen. Dabei fallen zwei Dinge vornehmlich auf. Zum einem sehen wir, wie gegen Ende 2016 die Varianz der Ausprägungen zunimmt. Dies hängt mit der Anzahl der Zählstationen zusammen. Die früheste Zählstation an der Renzstraße ging 2014 in Betrieb. Weitere folgten erst 2016. Je mehr Zählstationen mit unterschiedlichen Verkehrsströmen messen, desto höher ist die tägliche Varianz in den Zähldaten. Weiterhin fallen einige wenige weiße Stellen auf. Dies hängt mit unvollständigen Wetterdaten zusammen. Die Menge fehlender Daten fällt hier aber in einen vertretbaren Rahmen, wobei fehlende Tage nicht weiter mit wichtigen Variablen korrelieren.

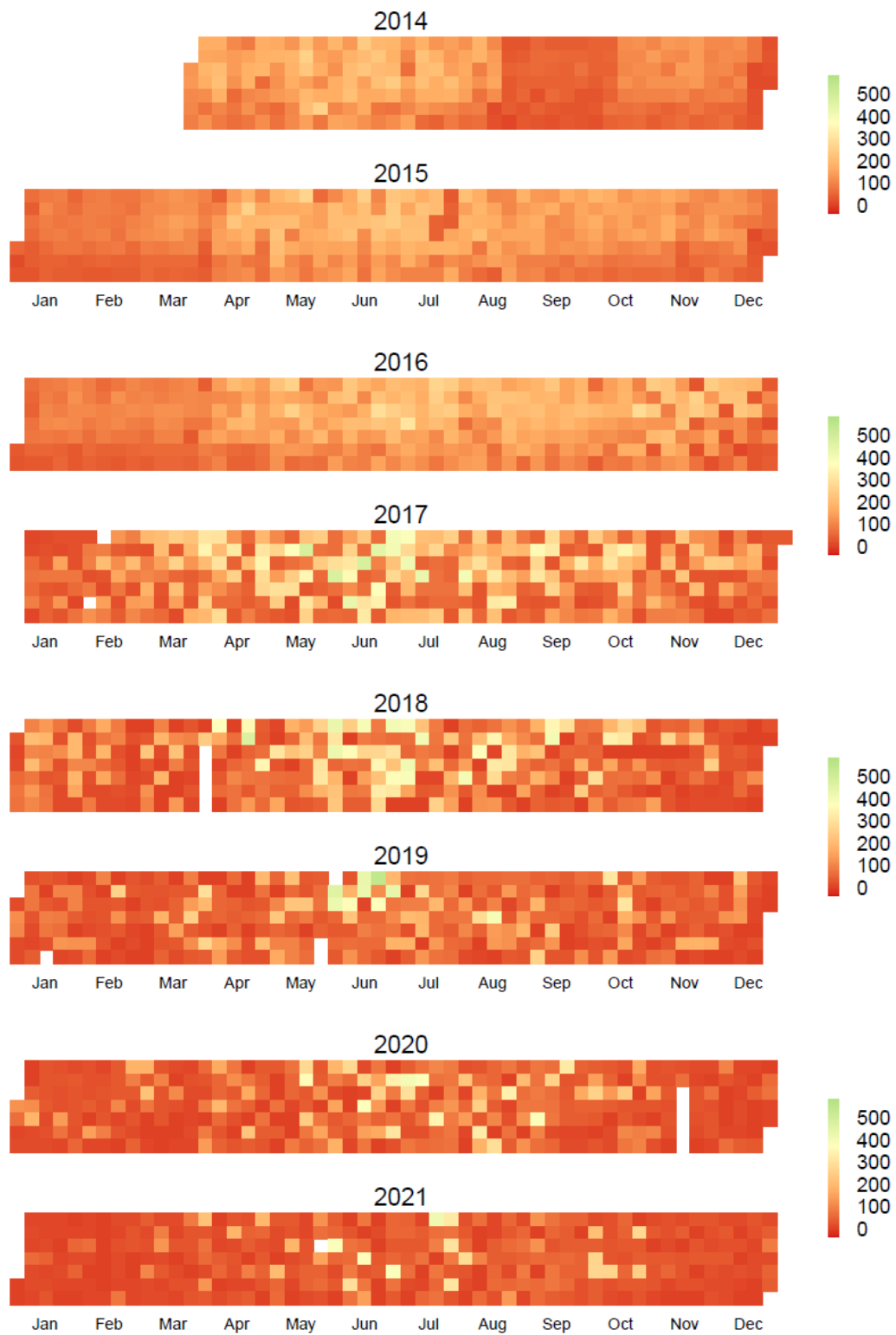
Einen Überblick über die acht Zählstationen in Mannheim ist in Abbildung 3.2 zu sehen, wobei zur Erstellung der Graphik ein R Paket genutzt wurde nach Kahle and Wickham (2013). Am Farbgradienten ist zu sehen, wie viele Fahrradfahrer pro Stunde die Zählstationen passieren. Zudem ist der Name der Zählstation eingeblendet zudem das Jahr, seitdem diese Station installiert ist.

Weitere Variablen wurden aus dem Datum erzeugt. So befindet sich im Datensatz eine binäre Dummy Variable für das Wochenende und eine für Sommermonate. Weiterhin nutzen wir Jahr, Monat, Tag und Stunde als Faktorvariablen, wobei wir nächtliche Stunden zwischen 22 und 5 Uhr ausgeschlossen haben.

Zu Abweichungen des Fahrradverkehrs konnte es während der Corona Krise kommen, wo Teile des öffentlichen Lebens still standen und damit auch das Verkehrsaufkommen. Deshalb beinhaltet der Datensatz zwei Variablen, eine Dummy Variable, die vor Ausbruch von Corona null ist und eine Dummy Variable für den Zeitraum der Kontaktbeschränkungen.

Tab. 3.1: Datensatz Übersicht

Name	Min	Durchschnitt	Max	Stand.abw.	Korrelation
Zählstand	1	108.35	955	120.21	1
Distanz z Uni	208.36	1122.51	1875.72	583.35	0.26
Wochenende	0	0.28	1	0.45	-0.21
Sommer	0	0.48	1	0.5	0.15
Feiertag BW	0	0.03	1	0.18	-0.09
Feiertag RP	0	0.03	1	0.18	-0.08
Schulferien BW	0	0.32	1	0.47	0.04
Semesterferien	0	0.48	1	0.5	-0.03
Niederschlag	0	0.07	30.4	0.49	-0.03
Temperatur	-13	12.67	39.4	8.33	0.24
Wind	0.1	3	14.4	1.65	-0.02
Rel. Feuchte	14	69.24	100	20.3	-0.23
Sonne	0	19.22	60	24.82	0.17
Bedeckung	0	5.72	8	3.14	-0.08
Corona	0	0.42	1	0.49	-0.16
Kontaktbeschr	0	0.11	1	0.32	-0.09

Calendar Heat Map of bicycle traffic in Mannheim*Fig. 3.1: Ausprägungen des Fahrradaufkommens*

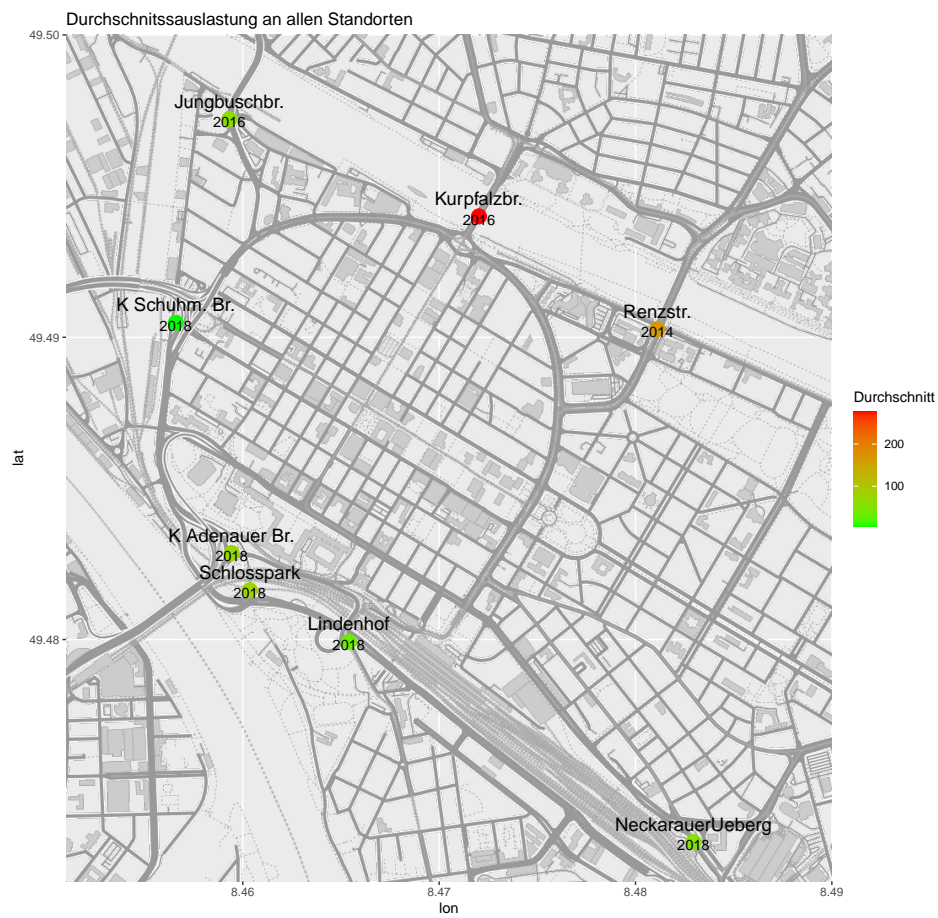


Fig. 3.2: Zählstationen im Überblick

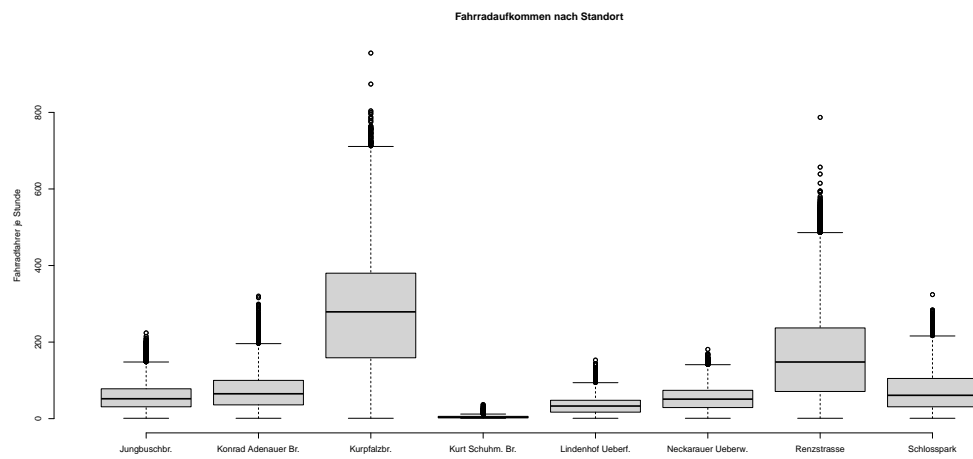


Fig. 3.3: Boxplots zu den Zählstationen

4. DESKRIPTIVE ANALYSE

4.1 *Regressionsanalyse*

Vorerst sind die gesammelten Daten trotz des Umfangs recht nutzlos, da sie allein wenig Antworten geben. Durchschnitte und Standardabweichungen einzelner Variablen, wie der Temperatur oder des Fahrrad Aufkommens, geben durchaus Aufschluss nie aber in einem größeren Zusammenhang. Doch was eigentlich von Interesse ist was den Fahrradverkehr beeinflusst. Dies mündet z.B. in der Frage welche Rolle der Regenfall beim Fahrradverkehr spielt, oder ob mehr Fahrradfahrer an Feiertagen auf Mannheims Straßen anzufinden sind. Dies muss nicht nur qualitativ, sondern auch quantitativ beantworten. Ein Mittel, um an diese Informationen zu kommen, ist die Regressionsanalyse. Diese versucht den Zusammenhang zu messen, in dem zwei Variablen zu einander stehen. So könnte man z.B. versuchen zu messen, wie sehr sich das Aufkommen an Fahrradfahrer unterscheidet in der Entfernung zur Universität Mannheim. Dargestellt ist dies in der Abbildung 4.1. Mithilfe der Regressionsanalyse kann man nun durch diese Punkte an den Positionen der acht Zählstationen jene Linie ziehen, die das Verhältnis mathematisch am besten repräsentiert, indem sie die quadratischen Abstände der Datenpunkte zur Linie selbst minimiert. Natürlich verhält sich nicht alles linear zu einander. Deswegen kann man auch auf quadratische oder kubische Terme zurückgreifen. Und wie man in der Darstellung sieht, beschreibt die kubische Relation scheinbar das Verkehrsaufkommen am besten. Doch ist es zu bezweifeln, dass allein eine Variable zu einer präzisen Vorhersage führt.

Nimmt man weitere Variablen in das Regressionsmodell auf, die den Fahrradverkehr erklären könnten, wie z.B. Wetterbeobachtungen oder Ferien- und

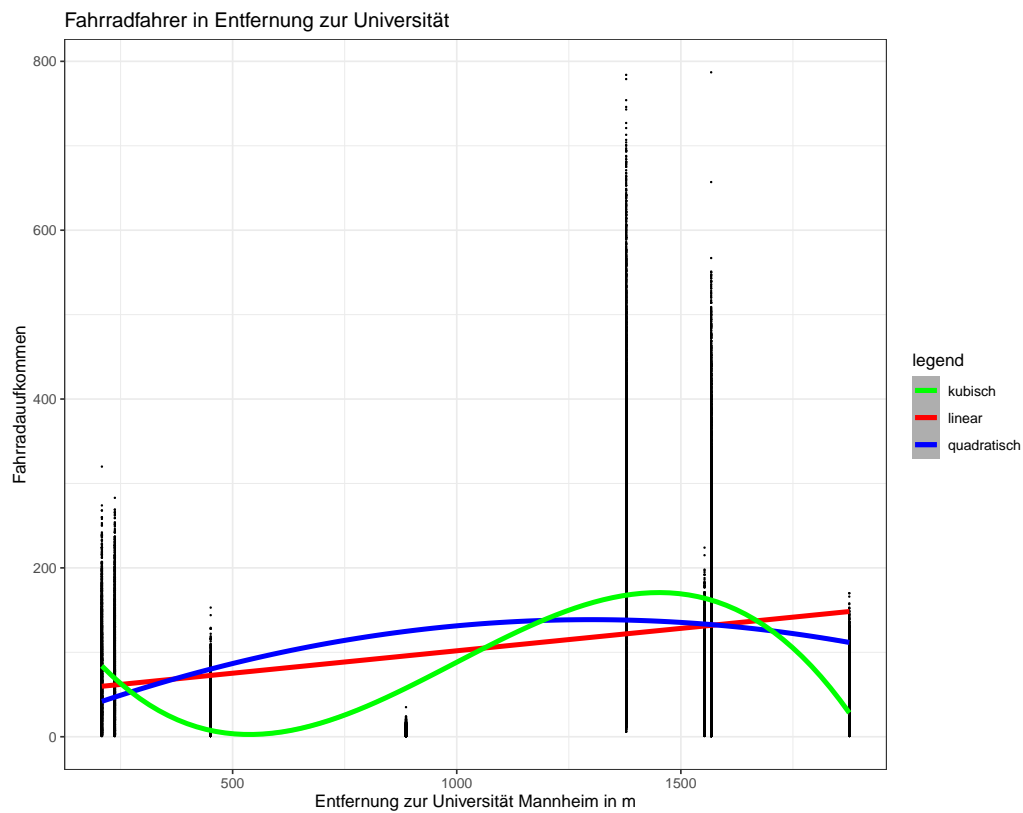


Fig. 4.1: Fahrräder in Entfernung zur Universität Mannheim
Graphisches Beispiel einer Regression

Feiertage, dann steigt die Genauigkeit (Fit) des Modells. Das heißt wir können präzisere Vorhersagen treffen, weil wir uns ein mehrdimensionales Modell schaffen. Dies nennt man dann auch die multiple Regressionsanalyse. Lässt man wiederum wichtige Variablen aus, dann kann das auch die Aussage über die Kausalität verzerren durch den Bias einer ausgelassenen Variable. Deswegen ist es wichtig, die richtigen und sinnigen Variablen in das Modell aufzunehmen. Im folgendem Modell sind Wetterdaten und Ferien- Feiertage mit inbegriffen. Zusätzlich verwenden wir eine Dummy Variable für die Sommer Monate und eine Dummy Variable für alle Zeiträume, die von Corona Kontaktbeschränkungen betroffen waren. Nicht direkt dargestellt in der Übersicht 4.1, sind Dummy Variablen, für die vollen Stunden, Wochentage, Monate und Jahre, sowie für die den Namen der Station als Faktor, damit das

Modell zwischen den verschiedenen Leveln der Zählstationen unterscheiden kann, denn der Verkehr an den einzelnen Stationen weicht stark voneinander ab, wie man auch in der Darstellung 3.3 sieht.

Tab. 4.1: Log-Lineares Regressionsmodell

	Koeffizient	Std Abw	t-Wert	Pr(> t)	
Leichter Nieselregen	-0.184	(0.015)	-12.12259	5.9389e-06	***
Starker Nieselregen	-0.249	(0.012)	-20.60850	1.5899e-07	***
Leichter Regen	-0.331	(0.021)	-15.84990	9.6482e-07	***
Moderater Regen	-0.353	(0.019)	-18.79747	2.9962e-07	***
Starker Regen	-0.373	(0.043)	-8.61747	5.6532e-05	***
Heftiger Regen	-0.209	(0.051)	-4.05542	4.8375e-03	**
Temperatur	0.036	(0.005)	7.63981	1.2217e-04	***
Temperatur ²	-0.0005	(0.000)	-9.29870	3.4479e-05	***
Wind	-0.041	(0.003)	-16.05586	8.8335e-07	***
rel. Feuchte	-0.004	(0.001)	-3.51258	9.8253e-03	**
Sonne	-0.0001	(0.000)	-1.37325	2.1204e-01	
Bedeckung	-0.009	(0.001)	-8.05733	8.7084e-05	***
FeiertagBW	-0.460	(0.081)	-5.69499	7.3917e-04	***
FeiertagRP	-0.414	(0.059)	-7.00091	2.1138e-04	***
SchulferienBW	-0.139	(0.020)	-6.84816	2.4239e-04	***
SemesterferienUM	-0.159	(0.031)	-5.19710	1.2570e-03	**
Sommer	0.082	(0.021)	3.99890	5.1972e-03	**
Kontaktbeschr	-0.133	(0.027)	-5.01609	1.5372e-03	**
Num.Obs.	410891		RMSE	0.62	
R2	0.818	R2 Adj.	0.818		
Fixed Effects	Standort	Stunde	Wochentag	Jahr	
Signifikanz:	*p<0.1;	**p<0.05;	***p<0.01		

Interessant sind für uns in der Tabelle 4.1 die Koeffizienten und die Signifikanzniveaus. Wobei wir berücksichtigen müssen, dass wir ein log-lineares Modell verwendet haben. Das heißt wir haben die abhängige Variable den Zählstand der Fahrradstation logarithmiert, um nicht lineare Einflüsse modellieren zu können. Die Signifikanzniveaus werden durch Sterne gekennzeichnet. Drei Sterne markieren das höchste Signifikanzniveau, kein Stern bedeu-

tet, dass der kausale Einfluss eines Koeffizienten statistisch insignifikant ist, also zu klein im Verhältnis zur Streuung der Daten. Die Vorzeichen der Koeffizienten zeigen uns, ob eine Variable dazu führt, dass weniger Leute Fahrrad fahren, wenn z.B. Regen fällt, oder weniger.

Den Regenfall haben wir kategorisiert, nach derselben Methode wie sie Wessel (2020) verwendet. Auffällig ist, dass der heftige Regen wohl einen kleineren negativen Einfluss hat als leichter Regen. Erklärbar wäre diese Beobachtung vielleicht dadurch, dass heftiger Regen oft plötzlich kommt und Fahrradfahrer überraschen kann. Auch der Mangel an Beobachtungen könnte dazu führen, denn heftiger Regen kommt selten vor, weshalb sein Signifikanzniveau auch geringer ist. Ansonsten sehen wir, dass zunehmender Regen einen zunehmend negativen Effekt hat. Für die Temperatur haben wir einen linearen und einen quadratischen Term in das Model aufgenommen, denn steigende Temperaturen führen nur so lange zu mehr Fahrradfahrern, solange es nicht zu heiß wird. Die quadratische Funktion kann dies sehr gut repräsentieren. Allerdings sind alle anderen Wetterbeobachtung längst nicht so einflussreich wie der Regen. Die Sonneneinstrahlung ist sogar insignifikant.

Einen recht starken Einfluss haben die Feiertage. Wobei wir unterschieden haben zwischen Baden Württemberg und Rheinland Pfalz. Hierbei sind die Effekte der Feiertage interessanterweise negativ. Dies gibt uns einen Hinweis darauf, dass in Mannheim ein utilitaristischer Fahrradverkehr vorherrscht. Damit ist ein Verkehr gemeint, der nicht der eigenen Unterhaltung und Erholung dient, sondern dazu dient einen Zweck zu erfüllen. In den meisten Fällen wird es sich hierbei also um Berufsverkehr handeln bzw. Verkehr hin zu den Bildungsstätten der Stadt. Darauf weisen auch die negativen Effekte für die Schulferien in Baden Württemberg und die Semesterferien der Universität Mannheim hin.

4.2 *Negativ Binomial Regression*

Die Zählstationen zählen immer in ganzen Schritten, da diese naturgemäß keine halben Fahrräder zählen können. Das bietet den Vorteil, dass wir eine so

genannte negativ binomial Regression verwenden können. Diese geht zurück auf Hausman et al. (1984), die in ihrem Modell auf die Poisson Verteilung aufbauen.

Tab. 4.2: Negativ Binomial Regression

	Koeffizient	Std Abw	t-Wert	Pr(> t)	
Leichter Nieselregen	-0.189047	(0.010061)	-18.790256	< 2.2e-16	***
Starker Nieselregen	-0.233913	(0.007726)	-30.274180	< 2.2e-16	***
Leichter Regen	-0.310315	(0.011664)	-26.605631	< 2.2e-16	***
Moderater Regen	-0.327817	(0.011601)	-28.257201	< 2.2e-16	***
Starker Regen	-0.305291	(0.024972)	-12.225408	< 2.2e-16	***
Heftiger Regen	-0.086290	(0.052804)	-1.634136	1.0223e-01	
Temperatur	0.035417	(0.002609)	13.576529	< 2.2e-16	***
Temperatur ²	-0.000425	(0.000045)	-9.433671	< 2.2e-16	***
Wind	-0.037234	(0.002592)	-14.364224	< 2.2e-16	***
rel. Feuchte	-0.002813	(0.000649)	-4.333943	1.4646e-05	***
Sonne	-0.000086	(0.000093)	-0.927537	3.5365e-01	
Bedeckung	-0.009645	(0.000888)	-10.865248	< 2.2e-16	***
FeiertagBW	-0.495062	(0.068356)	-7.242427	4.4073e-13	***
FeiertagRP	-0.247711	(0.031382)	-7.893448	2.9395e-15	***
SchulferienBW	-0.137121	(0.013632)	-10.058793	< 2.2e-16	***
SemesterferienUM	-0.127239	(0.005794)	-21.962002	< 2.2e-16	***
Sommer	0.080012	(0.008060)	9.926549	< 2.2e-16	***
Kontaktbeschr	-0.102099	(0.024805)	-4.116008	3.8549e-05	***
Num.Obs.	410891		RMSE	49.86	
psydo R2	0.164	psydo R2 Adj.	0.164		
Fixed Effects	Standort	Stunde	Wochentag	Jahr	
Signifikanz:	*p<0.1;	**p<0.05;	***p<0.01		

4.3 Utilitaristischer und Freizeitverkehr

Wie bereits bei Vorstellung des Standardmodells angeschnitten, geben die Variablen für Feier- und Ferientage einen Hinweis darauf, dass in Mannheim ein utilitaristischer Verkehr vorherrscht, also einen Fahrradverkehr hin zu den Arbeits- und Ausbildungsstätten der Stadt. Damit ist jedoch nicht un-

bedingt erwiesen, dass dies der Fall für alle Zählstationen ist. Wege die zu den Arbeits- und Ausbildungsstätten führen, können sehr wohl dem Berufsverkehr dienen, während andere Wege, weiter abseits, zur Erholung genutzt werden könnten. Dieser Verdacht lässt sich überprüfen mit einem Vergleich einzelner Regressionsmodelle über alle Fahrradzähler hinweg, die die notwendigen Variablen beinhalten.

Dies fasst die Tabelle 4.3 zusammen, die mithilfe einer R Bibliothek nach Hlavac (2022) erstellt worden ist. Die dargestellten Stationen sind der Reihe nach die Renzstraße, die Jungbuschbrücke, die Konrad Adenauer Brücke, die Kurt Schuhmacher Brücke, die Lindenhofüberführung, der Neckarauer Übergang und der Schlosspark. Auch hier sieht man, dass der utilitaristische Verkehr an allen Stationen überwiegt. Die Feiertage in Rheinland Pfalz spielen meist keine Rolle und sind insignifikant. Ansonsten führen Feiertage, Semesterferien und Wochenenden zu einer Abnahme des Fahrradverkehrs. Eine Ausnahme stellen die Schulferien dar. An der Renzstraße, der Konrad Adenauer Brücke, der Kurt Schuhmacher Brücke und am Schlosspark führen die Schulferien zu einer leichten Zunahme des Fahrradverkehrs. Dies könnte daran liegen, dass anders als die Studenten die Schüler mit ihren Schulen breiter über die Stadt verteilt sind, während die Zählstationen eher in der Nähe der Innenstadt und dem näherem Zentrum angesiedelt sind. Auch fahren Studenten innerhalb der Semesterferien häufiger nach Hause, und sind somit in der Stadt nicht mehr vorhanden. Den meisten Freizeitverkehr findet man insgesamt an der Kurt Schuhmacher Brücke. Zu mindestens sind hier die negativen Effekte der Koeffizienten am geringsten. Umgekehrt ist am Neckarauer Übergang der meiste utilitaristische Verkehr anzutreffen mit den größten negativen Effekten. Vergleicht man beide Stationen ist festzustellen, dass das Wetter den Freizeitverkehr stärker beeinträchtigt, während beim utilitaristischen Verkehr Radfahrer noch eher bei Regen, Wind und Wetter auf der Straße zu finden sind. Zusehen ist dies in der Tabelle 4.4.

Das Standardmodell kann man nun dazu verwenden, Prognosen zu machen. Für drei unterschiedliche Beispiel Szenarios findet man Vorhersagen in der Abbildung 7.2 im Anhang. Ausführlicher werden aber mögliche Vorhersagen und ihre Treffsicherheit im anschließenden Kapitel behandelt.

Tab. 4.3: Utilitaristischer und Freizeitverkehr der Zählstationen

	log(Zaehlstand)							
	Renzs	JBusch	Adenauer	Kurpfalz	Schuhm	Lindenh	NeckarÜb	Schlossp
FeiertagBW	−0.905 (0.057) ***	−0.779 (0.048) ***	−0.888 (0.070) ***	−1.033 (0.053) ***	−0.630 (0.068) ***	−1.127 (0.068) ***	−1.032 (0.064) ***	−0.742 (0.078) ***
FeiertagRP	−0.051 (0.059)	−0.174 (0.050) ***	−0.145 (0.073) **	−0.021 (0.056)	0.056 (0.071)	−0.090 (0.071)	0.018 (0.067)	−0.148 (0.082) *
SchulfBW	0.052 (0.007) ***	−0.054 (0.006) ***	0.126 (0.008) ***	0.006 (0.006)	0.152 (0.008) ***	−0.098 (0.010) ***	−0.055 (0.010) ***	0.026 (0.012) ***
Semesterf	−0.266 (0.007) ***	−0.059 (0.006) ***	−0.125 (0.008) ***	−0.098 (0.006) ***	−0.041 (0.008) ***	−0.135 (0.009) ***	−0.077 (0.008) ***	−0.122 (0.010) ***
Wochenende	−0.829 (0.007) ***	−0.804 (0.006) ***	−0.791 (0.008) ***	−0.691 (0.007) ***	−0.451 (0.008) ***	−0.883 (0.009) ***	−0.823 (0.008) ***	−0.776 (0.010) ***
Constant	5.128 (0.006) ***	4.133 (0.004) ***	4.272 (0.006) ***	5.655 (0.005) ***	1.321 (0.006) ***	3.647 (0.006) ***	4.038 (0.006) ***	4.246 (0.007) ***
Observations	90,269	59,209	44,793	60,359	44,649	36,788	37,438	37,386
R ²	0.151	0.265	0.203	0.196	0.082	0.265	0.243	0.158
Adjusted R ²	0.151	0.265	0.202	0.196	0.082	0.265	0.243	0.158

Notes: ***Significant at the 1 percent level.
 **Significant at the 5 percent level.
 *Significant at the 10 percent level.

Tab. 4.4: Freizeit- und Berufsverkehr im Vergleich

	Freizeit			Utilitarist.	
Leichter Nieselregen	-0.208830***	(0.025861)	<	-0.063944***	(0.013529)
Starker Nieselregen	-0.242996***	(0.044008)	<	-0.071652**	(0.023158)
Leichter Regen	-0.350410***	(0.051344)	<	-0.125781***	(0.025659)
Moderater Regen	-0.318279***	(0.070454)	<	-0.120595**	(0.037419)
Starker Regen	-0.269545*	(0.127235)	<	-0.128016	(0.073305)
Heftiger Regen	-0.055661	(0.205122)	>	-0.106119	(0.137777)
Temperatur	0.050447***	(0.002065)	>	0.011108***	(0.001199)
Temperatur ²	-0.000884***	(0.000061)	<	-0.000151***	(0.000040)
Wind	-0.034832***	(0.003017)	<	-0.012150***	(0.001781)
rel. Feuchte	-0.002215***	(0.000390)	<	-0.000288	(0.000243)
Sonne	-0.000013	(0.000284)	>	-0.000184	(0.000183)
Bedeckung	-0.007426***	(0.001721)	<	-0.002814*	(0.001122)
FeiertagBW	-0.140294	(0.111894)	>	-0.207367***	(0.053588)
FeiertagRP	-0.369612**	(0.115990)	<	-0.085082	(0.056197)
SchulferienBW	-0.065830***	(0.010924)	<	-0.065506***	(0.007072)
SemesterferienUM	-0.068875***	(0.010248)	<	-0.028236***	(0.006324)
Sommer	0.076692***	(0.014095)	>	0.029047***	(0.008655)
Kontaktbeschr	-0.084528***	(0.014799)	<	-0.036209***	(0.008175)
Observations	44,649			37,438	
Adjusted psydo R ²	0.093052			0.041806	

Fixed Effects: Standort: 8;Stunde: 18;Wochentag: 7;Jahr: 9

Notes: ***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

5. EIGENE FORSCHUNGSFRAGEN

Im vorangegangenen Kapitel haben wir verschiedene Regressionsmodelle vorgestellt zur Erklärung des Fahrradaufkommens an den Fahrradzählstationen in Mannheim. Nun kann man die Vorhersage Qualität dieser Modelle am Bestimmtheitsmaß messen. Dieses beschreibt wie hoch der Anteil der Streuung ist, der durch das Model erklärt werden kann. Je näher das Bestimmtheitsmaß an 1 ist, desto besser erklärt (fittet) das Model die Datenpunkte, die man beobachten kann. Jedoch ermittelt man damit nur die interne Validität, als das Maß dafür, wie gut das Modell die Daten beschreibt, die man schon kennt und mit denen man das Modell berechnet hat. Was ist aber, wenn man ein Modell entwickeln möchte, dass nicht nur Vorhersagen machen soll, für bereits bekannte Variablen, sondern auch für Variablen, die in der Zukunft liegen, oder vielleicht außerhalb des Rahmens der bereits bekannten Fahrradstationen?

Diese Fragen lassen sich aufteilen. Zunächst sollte man versuchen zeitliche Prognosen für die Zukunft an den Zählstationen zu machen. Im nächsten Schritt könnte man übergreifende Vorhersagen machen, die nicht nur an den Zählstationen selbst gemacht werden, sondern an beliebigen Straßen, womit eine Topographie des Fahrradverkehr berechnet werden könnte.

5.1 *Model Selektion*

Allgemein muss der Datensatz aufgeteilt werden, in ein Trainings und Test Set. Dies birgt den Vorteil, dass sich nicht nur die interne sondern auch die externe Validität messen lassen kann, in dem wir das Model nur mit dem Trainings Set berechnen und mit dem Test Datensatz ein zusätzliches Be-

stimmtheitsmaß des Models berechnen anhand der Vorhersagen des Modells im Test Set. Als Bestimmtheitsmaß nutzen wir den adjusted R^2 Wert. Die Trainings und Test Sets werden zufällig aus dem bestehenden Datensatz gezogen bei einem Split von 80 zu 20, sodass das Test Set 20 % des ursprünglichen Datensatzes umfasst.

Eine naheliegende Vorgehensweise wäre es, verschiedene Modelle zu erstellen, die verschiedene Variablen beinhalten und diese Modelle zu validieren, um im Vergleich das beste Modell für seine Vorhersagen zu nehmen. Natürlich könnte man auch mittels Stepwise Selection dieses Modell erstellen. Dies ist ein maschineller Prozess, der das Modell Durchgang für Durchgang mit den Koeffizienten füllt, die die größte Validität mit sich bringen. Dieser Prozess ist jedoch auch sehr rechenintensiv, vor allem bei einem Datensatz dieser Größe. Stattdessen greift diese Hausarbeit auf 30 manuell erstellte Modelle zurück, in denen nach und nach mehr Koeffizienten hinzugefügt werden. Es beginnt mit den Variablen zum Wetter. Die ersten Modelle verwenden den Niederschlag als Zahl, spätere verwenden die Kategorisierung des Niederschlags. Danach kommen die restlichen Variablen hinzu, wobei das Jahr nicht mehr faktorisiert wird, da sonst der Effekt für zukünftige Jahre verloren geht. Statt des faktorisierten Standorts verwendet das Modell Lageparameter wie den Längengrad oder die Entfernung zur Uni. Dann fügen wir quadratische und kubische Effekte hinzu und zum Schluss berechnen wir drei letzte Modelle mit dem faktorisierten Standort, als Lageparameter. Um die Modelle zu vergleichen schauen wir das Bestimmtheitsmaß der Modelle im Trainings und Test Datensatz. Dabei kommt die Graphik 5.1 zustande.

Auf der Y-Achse der Graphik 5.1 abgetragen ist die Höhe des Bestimmtheitsmaßes. Auf der X Achse erscheint die Anzahl der Koeffizienten. Mit jedem weiteren Model wurden weitere Variablen hinzugefügt, angefangen mit den Wetter Variablen, Feiertage, Ferien, Zeitliche Variablen, wie das Jahr, Sommer oder Tageszeit bis hin zur geographischen Lage. Wir sehen, dass das Bestimmtheitsmaß, also die Genauigkeit der Vorhersage des Modells, zunimmt je mehr Variablen hinzugefügt werden bis zu einem Punkt, an dem das Bestimmtheitsmaß im Trainingsdatensatz zwar nochmals bedeutend zunimmt, im Test Datensatz aber abnimmt. Hier geraten wir in das so genannte

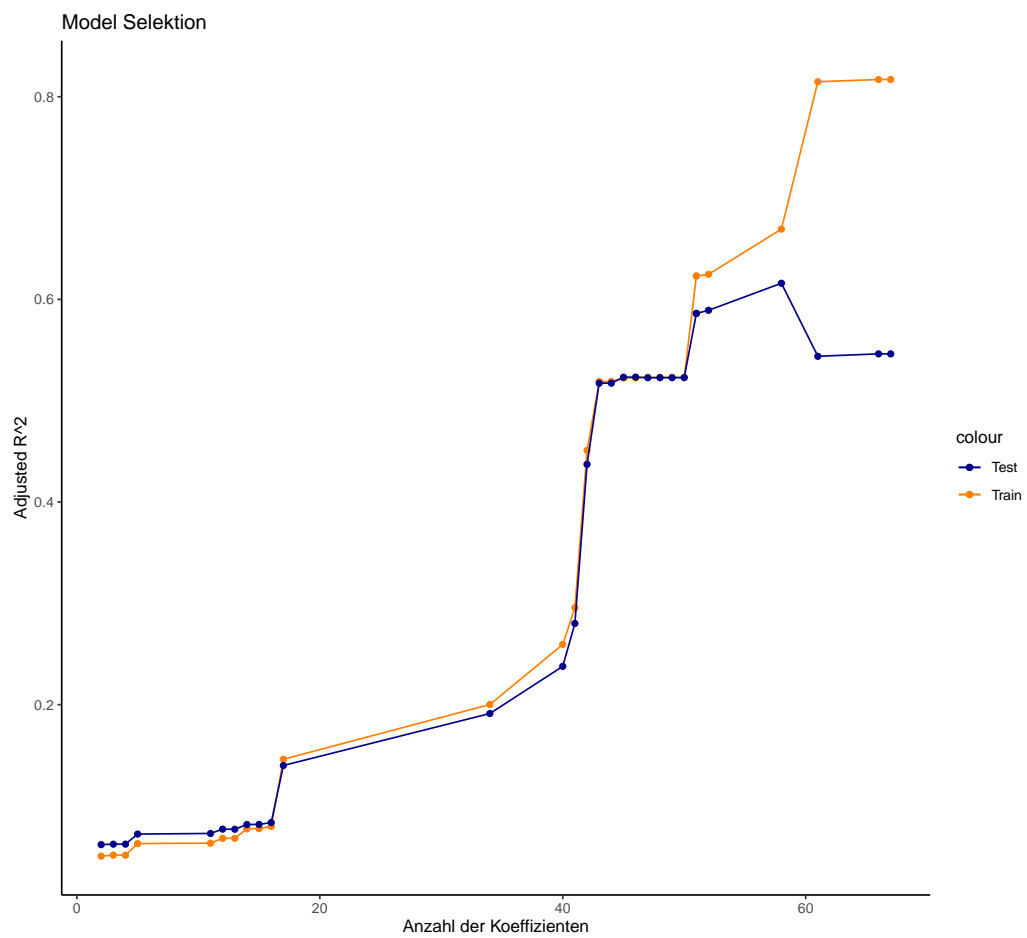


Fig. 5.1: Model Selektion

Overfitting. Das heißt, das Modell, das hier entwickelt wurde, kann zwar die Daten, mit denen es trainiert wurde, extrem gut erklären, sobald wir aber mithilfe des Modells Vorhersagen treffen wollen, auf Grundlage neuer Daten, ist die Genauigkeit der Vorhersagen schlechter, als sie sein dürfte. Das muss verhindert werden. Denn ist ein Modell von Overfitting betroffen, sind in den Variablen Zusammenhänge zu sehen, die sich nicht replizieren lassen.

Was löst in unserem Fall das Overfitting aus? Dazu vergleicht man den Regressions Output zweier Modelle in der Tabelle 5.1. Das zweite Modell ist jenes, in dem erstmals Overfitting auftritt. Das erste Modell ist das davor, in dem noch kein Overfitting auftritt.

Tab. 5.1: Modell Vergleich

	<i>Dependent variable:</i>	
	log(Zaehlstand)	
	(1)	(2)
Temperatur	0.028*** (0.001)	0.028*** (0.001)
Temperatur ²	0.0004*** (0.0001)	0.0005*** (0.0001)
Temperatur ³	−0.00002*** (0.00000)	−0.00002*** (0.00000)
Leichter Nieselregen	−0.188*** (0.007)	−0.187*** (0.005)
Starker Nieselregen	−0.269*** (0.012)	−0.263*** (0.009)
Leichter Regen	−0.305*** (0.013)	−0.307*** (0.010)
Moderater Regen	−0.356*** (0.020)	−0.353*** (0.015)
Starker Regen	−0.384*** (0.043)	−0.361*** (0.032)
Heftiger Regen	−0.270*** (0.078)	−0.232*** (0.058)
Wind	−0.066*** (0.006)	−0.070*** (0.005)
Wind ²	0.008*** (0.002)	0.008*** (0.001)
Wind ³	−0.001*** (0.0001)	−0.001*** (0.0001)
Rel. Feuchte	0.0003 (0.002)	−0.001 (0.002)
Rel. Feuchte ²	−0.0001*** (0.00003)	−0.0001*** (0.00003)
Rel. Feuchte ³	0.00000*** (0.00000)	0.00000*** (0.00000)
Bedeckung	0.003 (0.006)	0.002 (0.005)

Bedeckung ²	−0.003 (0.002)	−0.003* (0.001)
Bedeckung ³	0.0001 (0.0001)	0.0002* (0.0001)
SemesterferienUM	−0.129*** (0.003)	−0.142*** (0.003)
SchulferienBW	−0.156*** (0.004)	−0.145*** (0.003)
FeiertagBW	−0.388*** (0.026)	−0.383*** (0.019)
FeiertagRP	0.062 (0.162)	0.076 (0.121)
FeiertagBW*RP	−0.553*** (0.164)	−0.568*** (0.123)
Sommer	0.053*** (0.005)	0.047*** (0.004)
Längengrad	2,939,930.000*** (14,714.250)	
Breitengrad	493,297.900*** (2,540.769)	
LängengxBreiteng	−58,170.540*** (299.637)	
Längengrad ³	−281.543*** (1.343)	
Konrad-Adenauer-Br.		0.182*** (0.004)
Kurpfalzbrücke		1.549*** (0.004)
K. Schumacher-Br.		−2.612*** (0.004)
Lindenhofüberf.		−0.509*** (0.005)
Neckarauer Überg.		−0.042*** (0.005)
Renzstraße		0.863*** (0.004)
Schlosspark Lind.		0.178*** (0.005)
uniMA_dist	−0.013*** (0.0001)	
Corona	0.015** (0.006)	−0.021*** (0.005)
Kontaktbeschr	−0.122*** (0.005)	−0.119*** (0.004)
Jahr	24,671.730*** (2,362.121)	−5,005.576*** (1,768.591)
Jahr ²	−12.216*** (1.170)	2.492*** (0.876)
Jahr ³	0.002*** (0.0002)	−0.0004*** (0.0001)
Constant	−41,369,101.000*** (1,588,755.000)	3,351,009.000*** (1,189,707.000)
Observations	328,715	328,715
R ²	0.669	0.815
Adjusted R ²	0.669	0.815

Note:

*p<0.1; **p<0.05; ***p<0.01

Fixed Effects:

Stunde, Wochentag

Beide Modelle unterscheiden sich im Wesentlichen in der Wahl der Fixed Effects. Fixed Effects sind all jene Schätzer, die über die Zeit nicht variieren, also fix sind und nur zwischen den Entitäten variieren (Stock and Watson (2015)). In diesem konkreten Fall heißt das, ein Fixed Effect ist eine Variable die zeitlich konstant bleibt, sich aber zwischen den Fahrradzahlstationen unterscheidet. Das dient dazu, die Unterschiede zwischen diesen Stationen zu messen und im Modell mit aufzunehmen. Denn fällt der Blick auf die Graphik 3.3, sehen wir, dass die Stationen untereinander recht stark variieren. So herrscht auf der Kurpfalzbrücke, die von der Innenstadt über den Neckar führt, grundsätzlich mehr Radverkehr. Deswegen lässt sich z.B. eine Dummy Variable in das Modell aufnehmen, die 1 ist, wenn ein Beobachtungswert von der Kurpfalzbrücke stammte und 0 ist, wenn der Beobachtungswert von einer anderen Station stammte, dann erklärt diese Variable den Unterschied zwischen der Kurpfalzbrücke und den anderen Stationen im Modell, ohne den ursächlichen Grund dafür zu kennen. Dies erhöht aber die Treffsicherheit (Fit) des Modells, denn nun können die anderen Variablen ihren Einfluss präziser schätzen.

Eine solche Dummy Variable können wir für sieben von acht Stationen in das Modell aufnehmen. Bei der achten reicht es, wenn alle anderen Dummy Variablen 0 sind. Andernfalls bekämen wir ein multikollineares Modell, also ein eines in dem mindestens zwei Variablen linear abhängig von einander sind. Genau auf diese Weise sind wir im zweiten Modell vorgegangen. Dieses beinhaltet eine Dummy Variable für die Kurpfalzbrücke, die Lindenhofüberführung, den Neckarauer Übergang, die Renzstraße und dem Schloßpark/Lindenhof. Die übrige Station an der Kurt Schuhmacher Brücke ergibt sich, wie bereits erklärt, wenn alle anderen 7 Dummy Variablen 0 sind. Diese Vorgehensweise führt zu einem Bestimmtheitsmaß von 81,44 %, was sich gut anhört, wie aber bereits gezeigt, größtenteils auf Overfitting zurückzuführen ist.

Möglicherweise hängt dies damit zusammen, dass die Stations Dummy Variablen zwar Unterschiede erkennen, aber keinen ursächlichen Grund dafür finden. Wenn man aber Lageparameter als Fixed Effekt nutzt, hät-

te man einen möglichen ursächlichen Grund für den Unterschied, nämlich die Lage. Da Lageparameter zeitlich auch konstant sind, lassen sich diese ebenfalls sehr gut als Fixed Effect nutzen. Deswegen befinden sich im ersten Modell der Längengrad, sowie der quadrierten Längengrad, der Breitengrad, Längen- und Breitengrad miteinander multipliziert (Intersektion) und der Abstand zur Universität Mannheim in „uniMA dist“. Auf den quadrierten Wert des Breitengrades muss verzichtet werden, weil dies wiederum zu Multikollinearität führen würde und im Regressionsoutput ein Fehler auftauchen würde, wie sich zeigte. Auch führt eine Kombination aus Lageparametern und Standort Dummy Variablen zu Fehlern im Regressionsoutput, weswegen wir beide getrennt von einander nutzen.

Obwohl das Modell mit den Lageparametern ein kleineres Bestimmtheitsmaß von nur 66,8 % hat, muss gegen das Model mit Standort Variablen entschieden werden, wegen einer höheren externen Validität gemessen im Test Datensatz.

5.2 Fahrrad Topographie

Nun steht ein Modell zur Verfügung, dass auf Lagekoordinaten und den Abstand zur Universität Mannheim in Metern zurückgreift, um seine Vorhersagen zu berechnen. Was passiert, wenn wir in dieses Model Lageparameter eingeben, die gar nicht von den Zählstationen selbst stammten sondern von einem beliebigen Straßenpunkt?

Dazu kann ein Netz von Koordinaten über die Straßen Mannheims gelegt werden und mit einander so verbunden werden, dass ein Straßennetz erkennbar wird. Allerdings ist das modellierte Straßennetz nicht vollumfänglich, sondern beinhaltet nur eine persönliche Auswahl an Straßen, die meiner persönlichen Ortskenntnis nach, die meiste Relevanz haben. Eine genauere Methode gibt es ohne mehr Daten nicht. Beim Testen des Models fällt auf, dass das Modell, welches im vorherigen Kapitel ausgewählt worden ist, extreme Ausreißer vorhersagt, wie wir in Darstellung 5.2 zu sehen, vergleicht man dies mit der Verteilung des originalen Datensatzes.

Ein Grund für die extremen Ausreißer ist die Einführung der Intersekti-

Tab. 5.2: Verteilung der Vorhersagen

	Min	Median	Mean	Max
Datensatz	1	63	108.3	955
Vorhersagen von Model 21	12.5	101.2	123.1	509
Vorhersagen von Model 22	12.5	101.9	123.8	507.6
Vorhersagen von Model 23	12.5	101.3	123.3	509.6
Vorhersagen von Model 24	0	318.1	22604.4	514502.8
Vorhersagen von Model 25	0	318.1	22604.4	514502.8
Vorhersagen von Model 26	0	285.5	19143.1	425675.6
Vorhersagen von Model 27	0	168	2943.41	36717.38

onsvariabel *Laengengrad* \times *Breitengrad*. Das ist jedoch nicht weiter dramatisch, denn ohne große Verluste bietet sich eines der vorherigen Modelle an, welches in der Übersicht 5.3 zu sehen ist. Wendet man genau dieses Modell auf das angelegte Straßennetz Mannheims an, dann kommt man zu der Vorhersage in der Graphik 5.2. Zudem ist eine animierte Vorhersage auf YouTube verfügbar (Link: https://www.youtube.com/watch?v=nqP-_xEMWok). Beiden Darstellungen kann man entnehmen, dass der Verkehr im Westen der Stadt stärker sei, vor allem im Nordwesten. Es gibt jedoch keine Möglichkeiten, diese Aussagen weiter zu prüfen, als durch bloße ungefähre subjektive Ortskenntnis des Verkehrs.

Zudem ist die berechnete Topographie sehr ungenau. Anlaufzentren wie der Bahnhof oder der Jungbusch sollten stärker hervor treten. Außerdem je weiter die Vorhersage vom Stadtzentrum entfernt ist, desto weniger ist sie verlässlich, denn die Beobachtungspunkte, auf denen das Modell zur Vorhersage aufbaut, liegen alle sehr nah am Zentrum. Mit nur acht Beobachtungspunkten auf ein räumliches Model zu schließen, ist zu gewagt, um daraus überzeugende Schlussfolgerungen zu ziehen. Vielmehr dienlich war die Topographie, um zu zeigen, wie weit Overfitting in den Modellen vorhanden war, denn auch wenn Model 24 bis 27 in der Darstellung 5.1 keinen Hinweis auf Overfitting gaben, zeigte sich in der Anwendung der Modelle außerhalb der geographischen Lageparameter des vorhandenen Datensatzes unrealistische Vorhersagen mit extremen Ausreißern.

Tab. 5.3: Finales Modell: Model 23

	<i>Dependent variable:</i>
	log(Zaehlstand)
Temperatur	0.034*** (0.001)
Temperatur ²	-0.0004*** (0.00003)
Leichter Nieselregen	-0.183*** (0.009)
Starker Nieselregen	-0.271*** (0.015)
Leichter Regen	-0.301*** (0.016)
Moderater Regen	-0.351*** (0.024)
Starker Regen	-0.347*** (0.051)
Heftiger Regen	-0.192** (0.093)
Wind	-0.031*** (0.004)
Wind ²	-0.001*** (0.0004)
Rel. Feuchte	-0.006*** (0.001)
Rel. Feuchte ²	0.00001** (0.00000)
Bedeckung	0.001 (0.003)
Bedeckung ²	-0.001*** (0.0004)
SemesterferienUM	-0.201*** (0.004)
SchulferienBW	-0.077*** (0.004)
FeiertagBW	-0.458*** (0.031)
FeiertagRP	-0.344* (0.195)
FeiertagBW*RP	-0.077 (0.197)
Sommer	0.045*** (0.006)
Längengrad	123.791*** (0.314)
Breitengrad	93.715*** (0.350)
uniMA_dist	-0.001*** (0.00001)
Corona	-0.250*** (0.006)
Kontaktbeschr	-0.089*** (0.006)
Jahr	0.040*** (0.002)
Constant	-5,761.069*** (20.485)
Observations	328,715
R ²	0.523
Adjusted R ²	0.523

Note: * p<0.1; ** p<0.05; *** p<0.01
Fixed Effects: Stunde, Wochentag



Fig. 5.2: Fahrrad Topographie Mannheims an einem vergangenen Tag

5.3 Vorhersagevergleich

Laut der Verwaltung der Stadt Mannheim war die Zählstation an der Renzstraße vom Dezember 2020 bis zum Dezember 2021 defekt. Mittels eines wissenschaftlichen Hochrechnungsverfahrens wurden diese fehlenden Daten geschätzt.

Aufgrund dessen wurden die Daten in diesem Zeitraum von dieser Station aus dem Datensatz ausgelassen, um die Modelle nicht zu verzerren. Ein abschließender Vergleich ist in der Darstellung 5.3 zu finden. Die blaue Linie

kennzeichnet der Vorhersage des Models dieser Hausarbeit und die rote Linie kennzeichnet die Vorhersage der Hochrechnung.

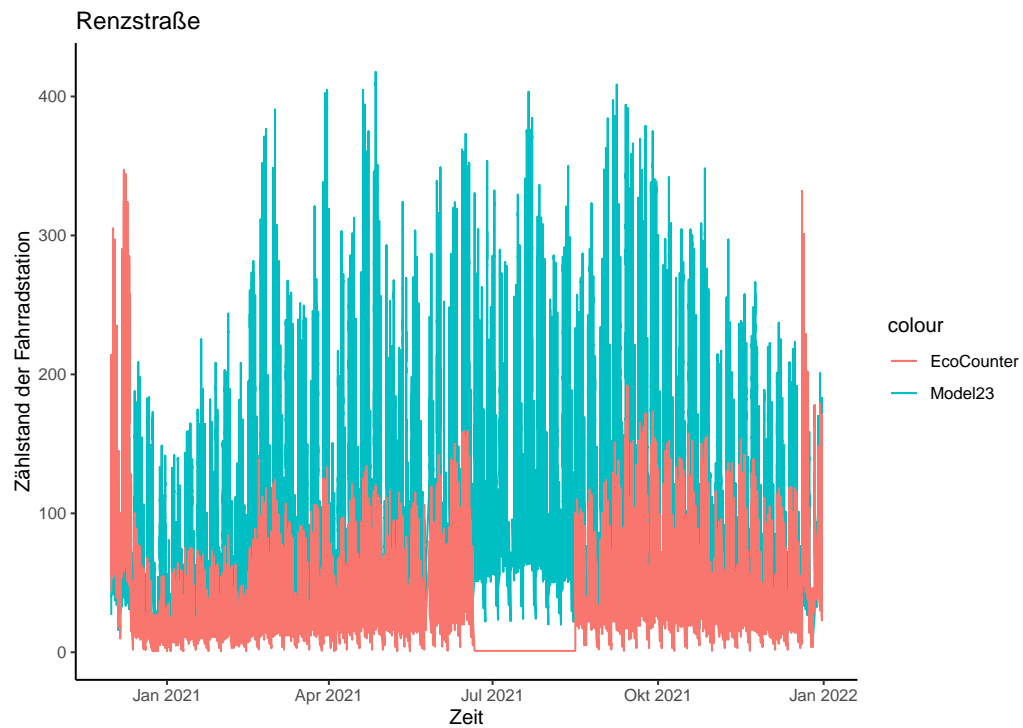


Fig. 5.3: Vergleich zur Hochrechnung an der Renzstraße 2021

6. FAZIT

Diese Hausarbeit bestand im Wesentlichen aus zwei Teilen. Der erste Teil beschäftigte sich mit der Frage, ob auf Grundlage von verschiedenen Parametern das Aufkommen von Fahrradfahrern an den beobachteten Zählstation erklärt werden kann.

Auf der Suche nach Daten, die man nutzen könnte, ließen sich Wetterdaten sowie Daten über Ferien- und Feiertage finden, die ein relativ solides Modell ergaben. Diesem Modell nach sind Niederschlag sowie Ferien- und Feiertage die wichtigsten Variablen bei der Entscheidung Fahrrad zu fahren. Zu erkennen ist auch, dass in Mannheim ein utilitaristischer Verkehr vorherrschend ist.

Der zweite Teil versuchte ein Modell zu berechnen, dass nicht nur dazu dienen sollte, beobachtetes Verhalten zu erklären, sondern es auch vorhersagen zu können. Die Frage war, ob sich das Modell extrapolieren lassen könnte, um eine Ansicht der Topographie des Fahrradverkehrs in Mannheim zu erstellen. Dazu mussten 30 verschiedene Modelle berechnet werden und bei der Auswahl des richtigen Modells fielen Hinweise zum Overfitting einzelner Modelle auf. Weitere Hinweise dazu traten bei der Berechnung der Topographie selbst auf, da das eigentliche Modell zu extremem Ausreißern führte. Erst durch eine weitere Reduzierung der Koeffizienten entstand eine schlüssige Vorhersage. Auch wenn der Aussagewert der berechneten Topographie angezweifelt werden kann, weil keine Daten verfügbar sind, um diese zu prüfen, war dieser Schritt insofern wertvoll, weil durch die Extrapolarisierung weiteres Overfitting der Modelle offen gelegt werden konnte.

So zeigt diese Hausarbeit auch, dass durch das Nutzen synthetischer zusätzlicher Datensätze die Robustheit statistischer Modelle erhöht werden kann. Eine robuste Schätzung der Topographie des Fahrradverkehrs könnte möglich

sein, wenn mehr geographische relevante Daten in das Model aufgenommen würden. Interessant wären z.B. auch die Entfernung zum nächst gelegenen Hauptbahnhof oder aber auch zum nächstgelegene Supermarkt, der Straßentypus auf dem die Zählstation liegt oder die Bevölkerungsdichte. Auch förderlich könnte es sein, diese Analyse auf mehrere Städte auszudehnen und die Ergebnisse in einem Modell zu bündeln.

7. ANHANG

Tab. 7.1: Wetterdaten als täglichen Niederschlag und Höchsttemperatur

	Standardmodell		taeglWetter	
Leichter Nieselregen	-0.183609***	(0.015146)		
Starker Nieselregen	-0.248702***	(0.012068)		
Leichter Regen	-0.331162***	(0.020894)		
Moderater Regen	-0.353010***	(0.018780)		
Starker Regen	-0.373162***	(0.043303)		
Heftiger Regen	-0.208576**	(0.051431)		
Täglicher Regen			-0.001410***	(0.000094)
Temperatur	0.036044***	(0.004718)		
Temperatur ²	-0.000482***	(0.000052)		
TagesHöchsttemp			0.034072***	(0.005517)
TagesHöchsttemp ²			-0.000353**	(0.000081)
Wind	-0.040923***	(0.002549)	< -0.035015***	(0.002136)
rel. Feuchte	-0.003787**	(0.001078)	> -0.004337**	(0.000948)
Sonne	-0.000118	(0.000086)	> 0.000179	(0.000115)
Bedeckung	-0.008657***	(0.001074)	< -0.000208	(0.000694)
FeiertagBW	-0.459770***	(0.080732)	< -0.444115**	(0.085660)
FeiertagRP	-0.414345***	(0.059184)	> -0.427126***	(0.065291)
SchulferienBW	-0.138772***	(0.020264)	< -0.138865***	(0.020478)
SemesterferienUM	-0.159223**	(0.030637)	< -0.150810**	(0.036192)
Sommer	0.082386**	(0.020602)	> 0.072395*	(0.022105)
Kontaktbeschr	-0.133188**	(0.026552)	> -0.139072**	(0.026333)
Observations	410,891		410,891	
Adjusted R ²	0.817542		0.817554	

Fixed Effects: Standort: 8;Stunde: 18;Wochentag: 7;Jahr: 9

Notes: ***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Tab. 7.2: Beispielhafte Vorhersagen

Beobachtungen:	(1)	(2)	(3)
Temperatur	28°C	4°C	12°C
Niederschlag	0 mm	3 mm	3 mm
Bedeckungsgrad	1/8	2/8	8/8
Windgeschwindigkeit	1 m/s	3 m/s	4 m/s
Luftfeuchtigkeit	75 %	85 %	71 %
Sonne	19.22	19.22	19.22
Feiertag BW	Nein	Ja	Nein
Feiertag RP	Nein	Ja	Nein
Schulferien BW	Nein	Ja	Nein
Semesterferien	Ja	Nein	Nein
Uhrzeit	16:00	11:00	15:00
Wochentag	Sonntag	Freitag	Mittwoch
Monat	August	November	Juni
Jahr	2018	2018	2018
Kontakteinschr	Nein	Nein	Nein
Vorhersagen:	356	32	162

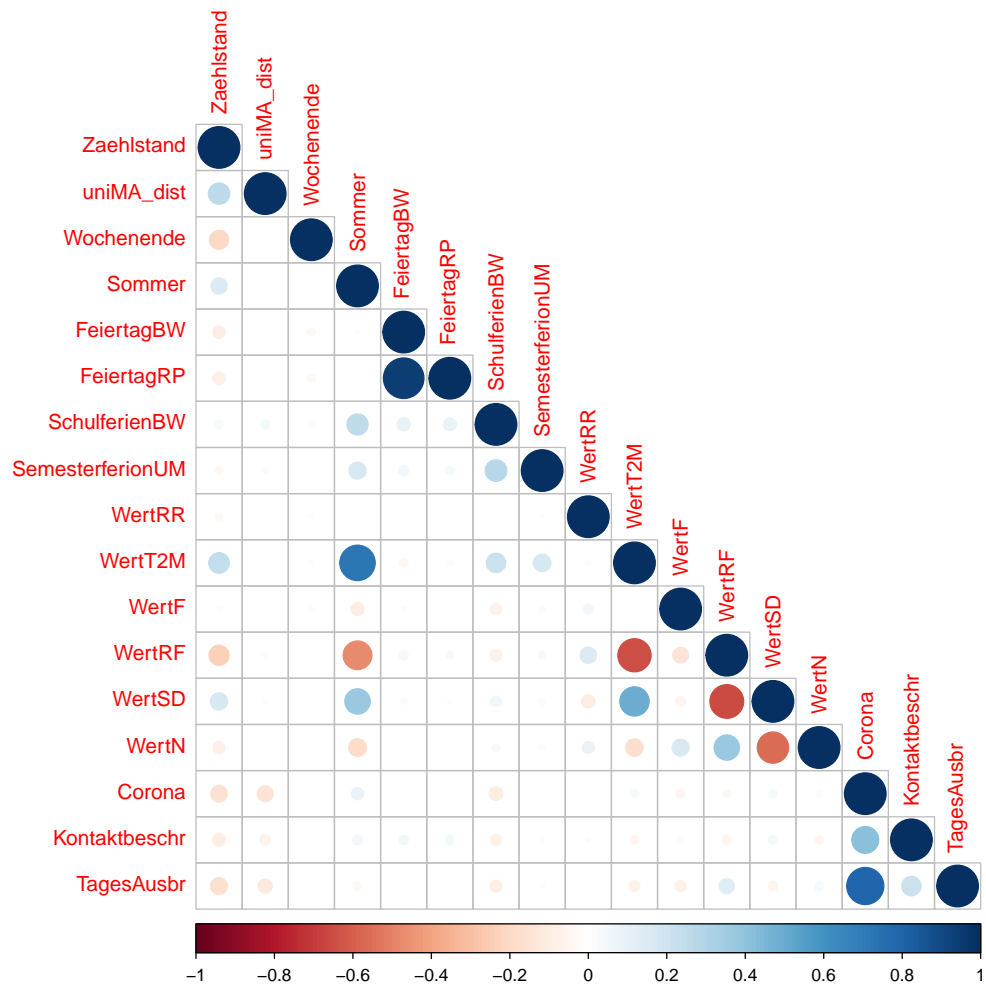


Fig. 7.1: Korrelations Plot

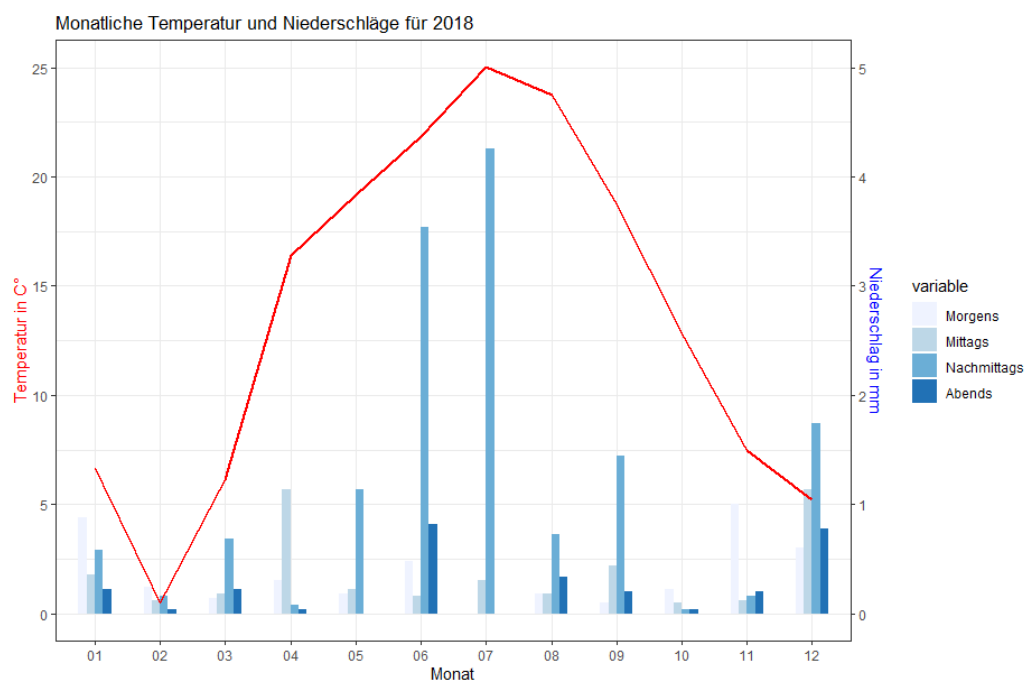


Fig. 7.2: Monatliche Durchschnittliche Temperaturen und maximale Niederschläge nach Tageszeiten 2018

Quelle: DWD, eigene Darstellung

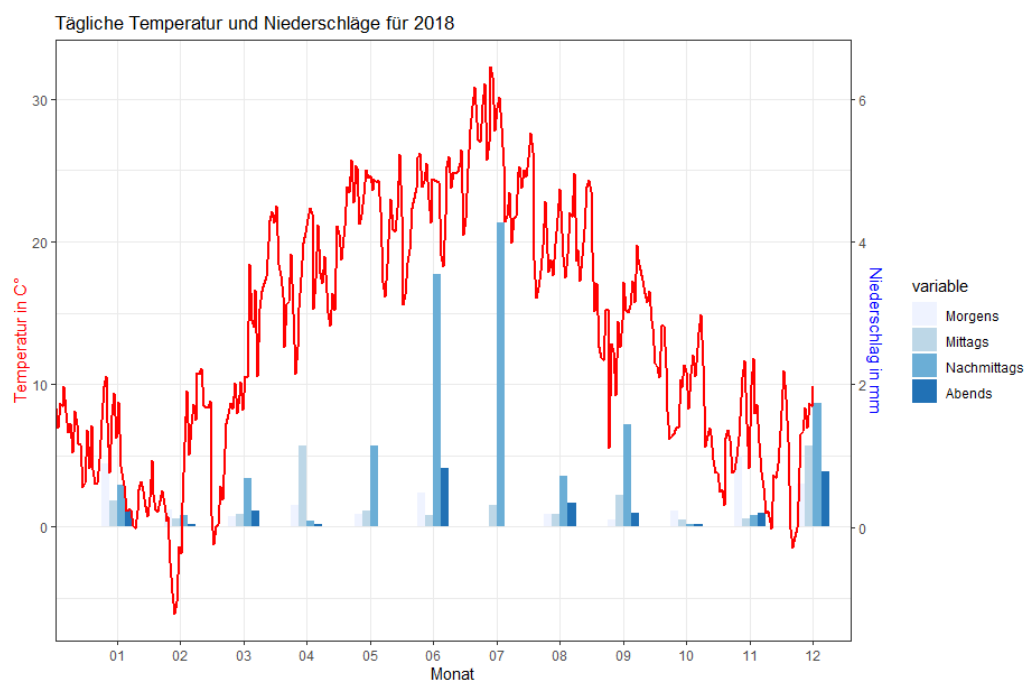


Fig. 7.3: Tägliche Durchschnitts Temperaturen und monatliche maximale Niederschläge nach Tageszeiten 2018

Quelle: DWD, eigene Darstellung

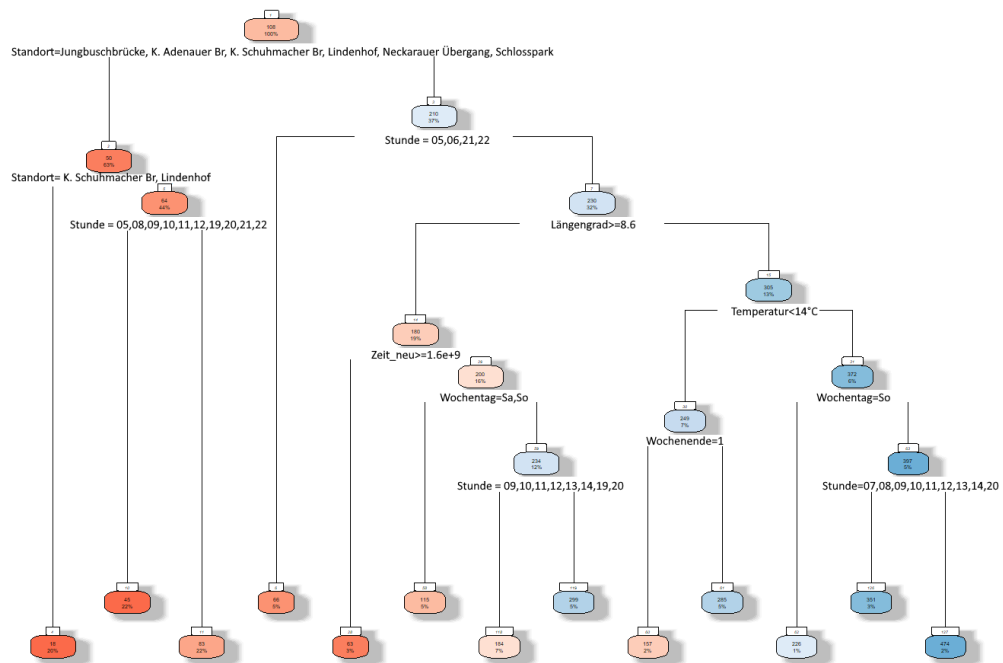


Fig. 7.4: Entscheidungsbaum

LITERATURVERZEICHNIS

- Hankey, S., Zhang, W., Le, H. T., Hystad, P., and James, P. (2021). Predicting bicycling and walking traffic using street view imagery and destination data. *Transportation Research Part D: Transport and Environment*, 90:102651.
- Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r and d relationship. *Econometrica*, 52(4):909–938.
- Hlavac, M. (2022). stargazer: Well-formatted regression and summary statistics tables. *R package version 5.2.3*.
- Hong J, McArthur DP, S. J. K. C. (2022). Did air pollution continue to affect bike share usage in seoul during the covid-19 pandemic? *J Transp Health*.
- Kahle, D. and Wickham, H. (2013). Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161.
- Stadt Mannheim (2010). https://www.mannheim.de/sites/default/files/2020-02/21-Pkt.-Programm_Platat.pdf 21-punkte programm für mehr radverkehr.
- Nobis, C. (2019). Mobilität in deutschland - analysen zum radverkehr und fußverkehr. *Studie von Infas, DLR, IVT, ind Infas 360 im Auftrag des Bundesministeriums für Verkehr und digitale Infrastruktur*, (70.905/90).
- Prati, G., Pietrantoni, L., and Fraboni, F. (2017). Using data mining techniques to predict the severity of bicycle crashes.

-
- Stock, J. and Watson, M. W. (2015). Introduction to econometrics. pages 403 – 407.
- Wessel, J. (2020). Using weather forecasts to forecast whether bikes are used. *Transportation Research Part A: Policy and Practice*, 138:537–559. Publication status: Published.
- Zhao, P., Li, S., Li, P., Liu, J., and Long, K. (2018). How does air pollution influence cycling behaviour? evidence from beijing. *Transportation Research Part D: Transport and Environment*, 63:826–838.