*Introduction company*

*Explanation assignment*

*Order processing*

*Description data*

*Analysis*

*Building the tool*

*Other project (On-Time)*

*Conclusion / Further research*

# Table of contents

Table of contents 2

1. Heineken 4

1.1 Breweries 4

1.2 Distribution units 4

2. The assignment 5

2.1 Events 5

2.2 Requirements of the tool 5

3. Order processing 7

3.1 Placing an order 7

3.1.1 The website 7

3.1.2 The call centre 7

3.1.3 The account managers 7

3.2 SAP 7

3.3 Order picking 8

3.4 Delivering the order 8

4. Description of the data 10

4.1 The dataset 10

4.1.1 Advantages of building your own dataset 10

4.1.2 Disadvantages of building your own dataset 10

4.2 Available data 11

4.2.1 Event forms 11

4.2.2 Order details in SAP 11

4.3 Usability of the data 12

4.4 Building the database 13

4.4.1 Factors 13

4.4.2 Product categories 15

4.4.3 The assortment table 17

4.5 Filling the database 17

5. Analysis 18

5.1 Influence of different factors 18

5.1.1 Expected number of visitors 18

5.1.2 Type of event 20

5.1.3 Indoor / Outdoor 21

5.1.4 Temperature 22

5.1.5 Precipitation 23

5.1.6 Conclusions 24

5.2 Influence of combinations 25

5.2.1 Temperature and type of event 26

5.2.2 Precipitation and type of event 27

5.2.3 Temperature and precipitation 28

5.2.4 Type of event and main categories 29

5.2.5 Temperature and main categories 30

5.2.6 Conclusions 33

6. Building the tool 34

7. Results 35

8. On-Time project 36

9. Conclusions and further research 37

Appendix 38

# 1. Heineken

Heineken International is the largest Dutch brewing company, and does not need an introduction for most people. Heineken is the third largest brewer in the world, with over 180 million hectolitres of beer produced per year. Once the British – South African brewing company SABMiller (known in the Netherlands for Grolsch and Bavaria) has been taken over by the world’s largest brewer AB InBev it will be the second largest brewing company.

Since the foundation in 1864, Heineken has grown to produce more than 250 different beers that are served in 192 countries. In addition to of course their own brand, they are known in the Netherlands for Amstel and Brand beer, as well as e.g. Jillz (apple cider) and Desperados (tequila flavoured beer). As a sponsor of the UEFA Champions League and organiser of the Holland Heineken House during the Olympic Games they have gained a lot of international publicity.

Heineken employs approximately 76.000 people internationally, owning more than 160 breweries in more than 70 countries. With the emergence of data analytics Heineken would like to know how they could improve different aspects of their supply chain. One of these aspects will be discussed in this paper.

## 1.1 Breweries

Text

## 1.2 Distribution units

Text

# 2. The assignment

Heineken has a lot of customers all over the world, and every customer is different. Their brands are served in many different types of restaurants, pubs and clubs, and their products are also available in supermarkets. Because these supply chains differ a lot from each other, they are part of different branches of the company.

Since the demand of supermarkets is quite consistent and therefore relatively easy to predict, the other branch is much more interesting in terms of data analysis. The consumption depends on how many customers visited the restaurant, pub or club, and not everybody drinks the same amount of beverages. During weekends and on holidays it is a lot busier, which influences the size of the order Heineken receives. To optimize the supply chain it is important for Heineken to be able to predict how much they should have in stock in order to supply each of their customers on time.

## 2.1 Events

Apart from the regular orders, customers can place additional orders for events they organise. As the organiser of an event you want to make sure that you order enough, so you will not be out of stock during the event and miss out on possible revenue. Therefore, in the case of an event the customer arranges with Heineken thatproducts that are not consumed can bereturned. If the amount of returned products is smaller than 15 per cent of the original order there will be no additional costs for the customer. However, when the amount of returned products is larger than 15 per cent, the customer has to pay for each extra product that he/she sends back to Heineken.

When products are returned there will also be costs for Heineken. Everything has to be collected, shipped back to thedistribution centre, sorted and put back in the warehouse. Therefore Heineken also benefits from having as few products as possible returned.

Currently the organisers determine the amount of products they order based on consumption during previous events, and if there are no previous events they have to guess how much they need for their event. They are assisted by one of Heineken’s account managers, who have a lot of experience with similar events, but the decisions are not data-driven. Therefore, Heineken would like to support their account managers by creating a tool to predict the consumption on events.

## 2.2 Requirements of the tool

Since the account managers have to use the tool eventually, they have to be able to quickly understand the functionality of the tool. Therefore, the dashboard has to be clear and intuitive. The users have to see at a glance where the input is, and what they have to give as input. Because the tool is going to be used to predict consumption and create orders based on those predictions, the users have to be able to convert or use the output as an order. What this means in practice will be explained in the chapter ‘Building the tool’.

Account managers do not have a lot of time when they want to make a prediction (they do not want to make the customer wait for a long time when they are creating an order), so it is also important that the tool is fast. Once all of the input fields are filled in, it should not take long before the output is generated. It is also possible that a customer changes his / her mind during the meeting with the account manager, in which case it has to be easy to compute new predictions for the event.

The third requirement is quite an obvious one, but nevertheless a very important one. Since the tool is going to be used to create orders, it directly influences the number of products supplied to the events. This means that the tool should always give an advice consisting of enough products, to prevent the organisers of the events from running out of stock. In order to achieve the goal of reducing the number of products that are sent back to Heineken, the tool should not predict a consumption that is too high either. Reliability is therefore another important requirement.

We can conclude that the main requirements of the tool are usability, performance in terms of speed and reliability. We will take these requirements into account when building the tool, and we will elaborate further on how we are going to do this in the chapter on ‘Building the tool’.

# 3. Order processing

There are a lot of steps between the customer placing an order and receiving that order. In order to understand the data and the importance of the tool in the order process we will explain the different steps of placing and processing orders.

## 3.1 Placing an order

There are three ways for customers to place an order: on the website, by calling the Heineken call centre (located in Houten, the Netherlands) or through their account manager (in the case of an order for an event). Each method is slightly different:

### 3.1.1 The website

When a customer visits the website, he / she can browse through the entire assortment that Heineken offers to their customers. The customers can log in with their own customer ID, and place their order. This order is then sent to Heineken and entered into SAP (see chapter X for more information on SAP).

### 3.1.2 The call centre

Customers can also place their order by telephone. The call centre in Houten receives every call, and processes the order. The employees in the call centre do not only take customer calls, they also make calls themselves. A lot of regular customers, such as restaurants and cafés, place their order on the same day every week. A call centre employee calls them one day before they want to receive the products and their order is entered into SAP.

### 3.1.3 The account managers

As mentioned before, the account managers assist the customers in creating the orders for events. They sit together with the customer and discuss which products the customer wants to order. Once they have decided what the order is going to be, the account manager fills in an event form with the details of the customer, the event and the order. The order is then entered into SAP.

## 3.2 SAP

SAP is the order processing system used by Heineken. In SAP all of the information about the stock is stored, as well as the information on the orders. As explained in the previous sections, (almost) every Heineken employee uses SAP. Since all of the supply chain information is stored here, it is a very useful source for data analysis. For example, the historical stock levels can be used to reduce future stock levels to save money. The historical order data can be used to predict future orders, as is the case in this project.

## 3.3 Order picking

In every distribution unit there is a large warehouse where the products are stored. Each different product has its own place on a shelf or on the floor, and has its own barcode. When an order has to be picked, the warehouse employees scan the order barcode to receive a list of products. They collect the products and scan the product barcode.



*Source:* [*http://www.ed.nl/economie/helmonds-heineken-voortaan-uit-oss-1.2139824*](http://www.ed.nl/economie/helmonds-heineken-voortaan-uit-oss-1.2139824)

This process minimizes the chances of someone accidentally forgetting a part of the order, since it is not possible to finish the order without scanning every product (unless, of course, a product is out of stock). Once they have finished an order, their hand terminal gives instructions on where to deliver the cart with the products. This way the truck drivers know where to find the carts with the orders they have to deliver the next morning.

## 3.4 Delivering the order

Each morning the truck drivers arrive at the distribution unit and receive a list with the customers they have to visit that day. They connect their truck to the dock they have been assigned, where the carts with their orders are placed. Before loading everything in their truck, they have to double-check every order, to make sure the order pickers did not forget something.



*Source: http://www.archiproducts.com/en/products/77787/loading-dock-kdr-kopron.html*

Once they have checked everything, they load the carts in their truck and start their route. Each customer has specified a time window, within which the products have to be delivered. The truck driver has to make sure that he / she delivers the order at the right time; otherwise the customer will likely file a complaint. More information about this can be found in chapter X (On-Time project).

# 4. Description of the data

In this chapter we will not only give a description of the data, but also explain how the data is stored and how we had to obtain the data.

## 4.1 The dataset

Since there wasn’t a dataset already available, we had to create our own dataset. There were some advantages, but of course also some disadvantages to this. Those advantages and disadvantages will be discussed shortly.

### 4.1.1 Advantages of building your own dataset

* *You are able to decide immediately which features you do, and which features you do not want to include in the dataset*

Instead of having to filter the data before you can start you analysis, you can decide beforehand whether you want to include features in the dataset or not. This saves time and prevents you from having to save a lot of different versions of the dataset (with different combinations of features).

* *You are able to determine the formatting of data*

When you enter the data into the dataset, you can make sure that it is in the correct format. For example, you can enter dates in the date format you prefer, instead of having to perform a transformation on that column afterwards.

* *You know exactly what the dataset looks like*

While this may seem obvious, it is still a big advantage. Instead of having to get familiar with the data first, you already know exactly what is in the dataset. This also gives a better overview of the possibilities for further analysis.

### 4.1.2 Disadvantages of building your own dataset

* *It takes a lot of time to build your own dataset*

If you have a ready-to-use dataset, you can immediately start exploring the data and make a method of approach based on the data that is provided. If you have to build your own dataset this is not the case. It takes time to decide what kinds of data you want / need, it takes time to gather the data and it takes time to enter the data into your dataset.

* *You have to decide what data you want to use before you start the analysis*

When you are analysing the data and you decide that you want to include certain features after all, you cannot just take them from the original dataset, since the features were left out in the first place. Therefore it is very important to determine the types of analysis you want to perform, so you can make sure all the data you need is added to the dataset.

* *There is a higher chance that mistakes are made*

Even though you can check the data you put into the dataset, a mistake is made easily. It requires a lot of carefulness to ensure that the data in the dataset is correct and reliable.

## 4.2 Available data

To create a dataset, we received the event forms of every event for the Amsterdam distribution unit in 2014. The event forms are digital files, although the account managers print an event form when they are meeting with a customer. Using these event forms we could look up the order in SAP to find the products that were actually delivered, and the products that were sent back.

### 4.2.1 Event forms

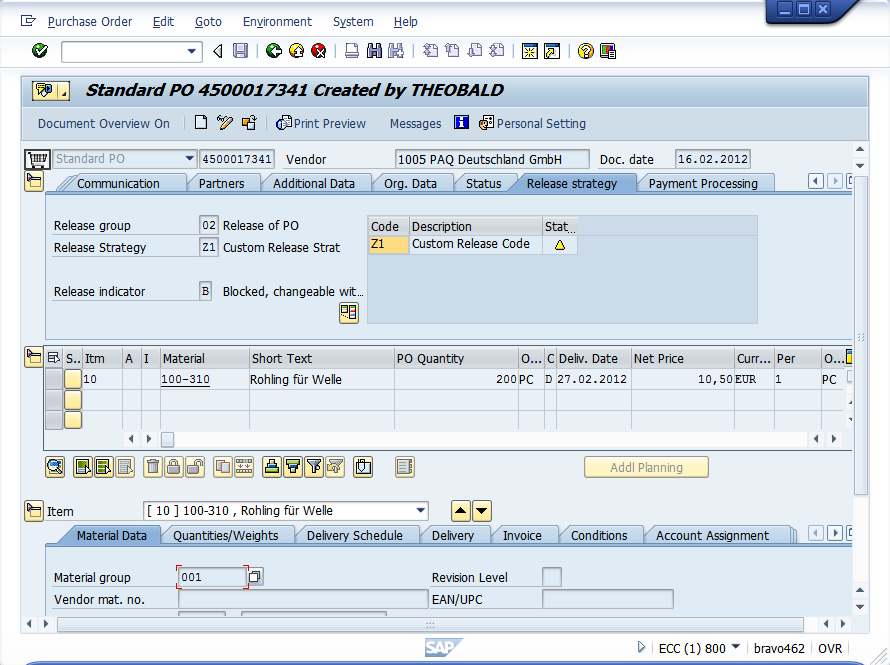
There were event forms for approximately 250 events, varying from big dance festivals to sporting events, and from rock concerts to parties in cafés. An event form contains all of the necessary information regarding the order for the event. Some important elements on the event form are:

* *Name of the event:* Naturally, the name of the event is included on the event form.
* *Ship-to number:* The identification number of the customer that placed the order.
* *Expected number of visitors:* The organiser and the account manager estimate the number of visitors of the event.
* *The order:* The details of the order are on a separate tab.

For an example of an event form, see appendix X.

### 4.2.2 Order details in SAP

For each event we looked up the order details in SAP, using the ship-to numbers on the event forms. In SAP every order has a separate number, which makes it easy to differentiate between different orders placed by the same customer. Since deliveries and returns are both entered as orders, there is a column indicating whether an order is a delivery or a return.



*Source: http://help.theobald-software.com/ERPConnect-Services-EN/default.aspx?pageid=bcs-purchase-order-release-in-sap*

When looking at a separate order, there are several important columns that we can use:

* *Product code:* Each product has its own unique product number, used to identify each product.
* *Amount ordered:* The amount of the product that has been ordered.
* *Product description:* A description of the product, including the brand, type and volume (e.g. Heineken cans 6x4x33cL)

This information will all be included in the dataset, because in order to predict the consumption at future events, we have to know how much has been consumed at previous events.

Unfortunately it is not possible to directly export an order from SAP, so we had to manually copy the contents of each order into the dataset. In the section on ‘Filling the database’ (chapter X) we will explain how we automated this as much as possible, to save time and to make it easier for future users to add new events to the event database.

## 4.3 Usability of the data

There was data available for approximately 250 events, but not every event was suitable to be included in the dataset. There are several reasons for an event not to be included in the dataset:

* *The event did not have a return order*

For some events such as fairs, especially events organised by cafés, there was no return order in SAP. When cafés organise an event and they still have products left after the event, they keep it and sell it later. Because of this it is not possible to know the exact consumption at the event, and therefore we cannot include these events in the dataset.

* *There is missing information*

On some event forms not all fields are filled in. Some information is crucial, such as the number of expected visitors at the event. If this information is not present, the event cannot be used and is therefore not included in the dataset.

* *The order is only half an order*

This category of orders is not completely useless, but needs to be treated differently. Not every customer orders everything at Heineken. Organisers may choose to order their beer at Heineken, but their wine and their soft drinks at another supplier. Since we cannot be sure that these customers order all products of one category (a category is e.g. beer, wine), the data of these events needs to be handled with care. These events can probably be used for analysis on consumption per category, but definitely not to predict the total consumption.

## 4.4 Building the database

Because of availability and usability, we chose to build the database in Excel. There was no server available to run the database on, so the tool and the database have to be ‘portable’. Therefore an Excel file with different tables is used.

With the usable events, we could start building the dataset. However, before entering the data of the events in a database, we first had to determine what kind of additional information we wanted to collect. Since some of the events are very different from each other, we cannot build a prediction model based on just the expected number of customers. Therefore a number of factors had to be specified, to be able to differentiate between different events.

### 4.4.1 Factors

* *Expected number of visitors*

This information is already available on the event forms, but it is still an important one to include in the list of factors. The number of visitors has a large influence on the total consumption at the event, so this is one of the key factors.

* *Type of event*

To determine the different types of events, and assign a type to each individual event, we made use of the experience of the account managers and other people involved in the event service. We did not want to create too little event categories, to prevent events that are too different from each other to fall in the same category, but we did not want to create too many categories either, because the information gain would be too little. Eventually we came up with the following five categories:

* *Village festivals:* Local festivities and fairs
* *House:* House- and dance parties and festivals (e.g. Amsterdam Open Air)
* *Jazz:* Jazz concerts and festivals (e.g. North Sea Jazz festival)
* *Rock:* Rock concerts and festivals (e.g. Oranjerock)
* *Sport:* Sporting events (e.g. Nederland – Wales)
* *Indoor / outdoor*

Another factor that has been included in the dataset is whether an event is indoors or outdoors. An event being indoors or outdoors has two important consequences:

* The influence of the weather is significantly less for an indoor event. For outdoor events the weather may influence the consumption pattern of visitors, because people prefer other drinks when it is cold than when it is warm. For indoor events this does not matter that much, because the temperature can be regulated and is usually approximately the same for all indoor events.
* The expected number of visitors is usually more accurate for indoor events. Most of the times you have to buy tickets for indoor events, while this is not always the case for events that are outdoors. Referring back to the previous consequence, the probability of visitors not showing up due to the weather is smaller for indoor events than for outdoor events.
* *Temperature in degrees Celsius*

The temperature on the day of the event is also added to the database. If an event lasts more than one day, the average temperature over the days is taken. As mentioned in the previous section, the temperature influences the consumption pattern of visitors. When it is warm, people may drink more water than when it is cold, while people may drink more red wine when it is cold than when it is warm.

* *Precipitation in millimetres*

The last factor that has been included is the precipitation (usually in the form of rainfall) in millimetres. It is possible that whether it is dry or it is raining influences the consumption behaviour of visitors. This will be investigated further in the next chapter ‘Analysis’.

In addition to the factors, there was another part of the database that had to be constructed. In SAP, the product description and the amounts of products are available, but this amount does not say anything about the product category and the number of litres. Therefore another table, containing the product category and the number of litres per product needs to be added to the database.

### 4.4.2 Product categories

To be able to predict more accurately than just the total consumption at events, the products are divided into different product categories. Since Heineken does not only supply beer but also many other products such as wine and soft drinks, there are six main categories, consisting of several subcategories:

* *Beer:*

Heineken is known for its beer, so naturally beer is a main category. However, next to their lager, Heineken supplies a lot of different types of beer:

* Cider
* Malt beer
* Lager
* Radler (beer and sparkling lemonade)
* Rosé beer
* Specialty beer
* Beach beer (e.g. Desperados, tequila flavoured beer)
* Wheat beer
* *Soft drinks:*

Heineken also supplies a lot of different types of soft drinks. Vrumona, a producer of soft drinks, is part of Heineken and produces a lot of the soft drinks that Heineken supplies. Apart from the brands that Vrumona produces, there are also a lot of ‘external’ brands in the assortment. The soft drinks subtypes in the database are:

* Bitter lemon
* Cassis
* Chocolate milk
* Coke
* Energy drinks
* Fernandes (a soft drink of Surinamese origin)
* Yoghurt drinks
* Ginger ale
* Ice tea
* Seven up
* Orange soda
* Tonic
* Fruit juice
* *Liquors:*

In the liquors category all the strong liquors are included. These are not produced by Heineken, but they are supplied to customers:

* Apfelkorn
* Cognac
* Gin
* Jenever
* Liqueur
* Mixed drinks
* Port
* Rum
* Sherry
* Tequila
* Vodka
* Whisky
* *Water:*

Water is a relatively simple main category, since there are not a lot of variations on water. It is important as a separate main category nevertheless, because it is non-alcoholic and significantly different from soft drinks. The three subcategories are:

* Uncarbonated water
* Sparkling water
* Vitamin water
* *Wine:*

Like the liquors, the wine Heineken supplies is not produced by Heineken itself, but customers can still order a large variety of wines from the following subcategories:

* Mulled wine
* Sparkling wine
* Red wine
* Rosé
* Vermouth
* White wine
* *Others:*

The final main category consists of other products, like glasses, plastic cups and coasters. These products are not used in the prediction model, but they are included so they can be used for ordering in the tool.

### 4.4.3 The assortment table

In the assortment table, information on all of the products is stored. The table contains a few columns with information to make entering data into the database easier, and a few columns that can be used for analysis and prediction. We will shortly discuss the database elements.

* *Product code:* The unique code that is assigned to a product. When an order is entered into the database, the product code can be looked up in the table to easily add the product to the order in the database.
* *Main category:* The main category the product belongs to.
* *Subcategory:* The subcategory the product belongs to.
* *Description:* A description of the product.
* *Units per package:* The number of units in each package. A crate of beer, for example, contains 24 bottles.
* *Volume per unit in litres:* The volume of one unit of the product. The beer bottles in the previous example contain 0,33 L.
* *Total volume per package:* The total volume of one package of the product, equal to the units per package times the volume per unit.

## 4.5 Filling the database

After determining and specifying the factors and product categories, the database was ready to be filled. For every event form order was looked up in SAP. The expected number of visitors was taken from the event form and filled in the table. The event type was not always filled in on the event form or sometimes the event type was not one of the specified event types in the tool. In that case the account manager responsible for that event was contacted to make sure that the correct event type was entered. This was also the case for the indoor / outdoor factor.

The temperature and the precipitation were obtained from (link X en link X). As mentioned in the section on the factors, the temperature and precipitation were averaged if the event lasted multiple days.

To make the input of a new order easier, a small VBA tool was created in which the user can paste the order copied from SAP. The tool then displays for every product in the order whether it is already included in the assortment table, and it adds the products that are included in the assortment table to the order in the database.

This is of course less efficient than exporting an order in one click, but in this case the possibilities were limited. This way the number of manual actions is minimized, so the database can be filled as soon as possible.

# 5. Analysis

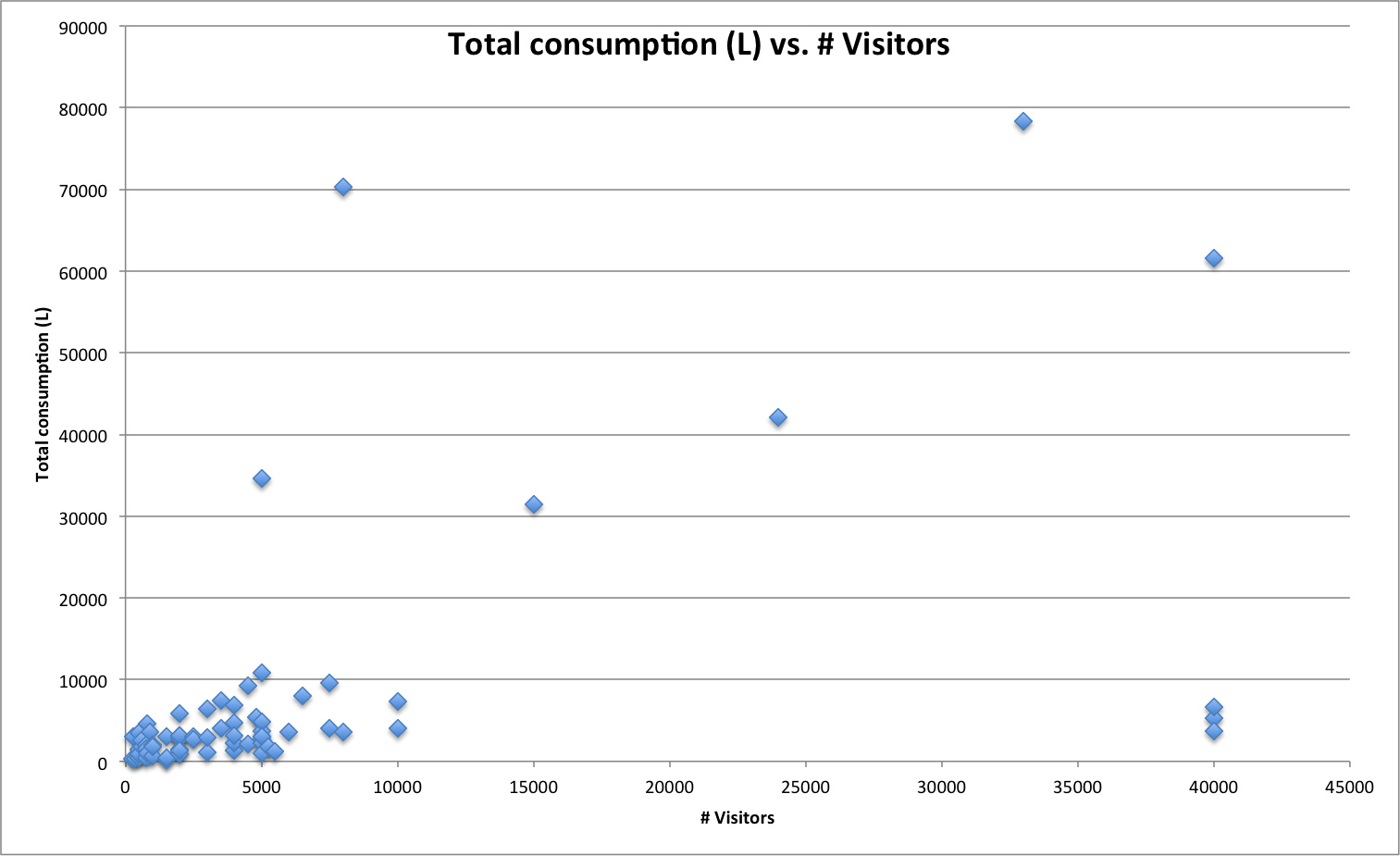
Before we can make a prediction model, we first have to analyse the influence of the different factors on the consumption. In the next sections we will discuss the influence of the different factors, and take a closer look at combinations of factors. Since consumption patterns are sensitive to trends, we will also investigate the consumption of different categories over time.

## 5.1 Influence of different factors

First we will take a look at the influence of the individual factors on the total consumption, and the consumption per main category. The results of this analysis can be used as input for a regression model for example. After analysing the individual factors, we will try combining factors to see whether they are correlated with the consumption.

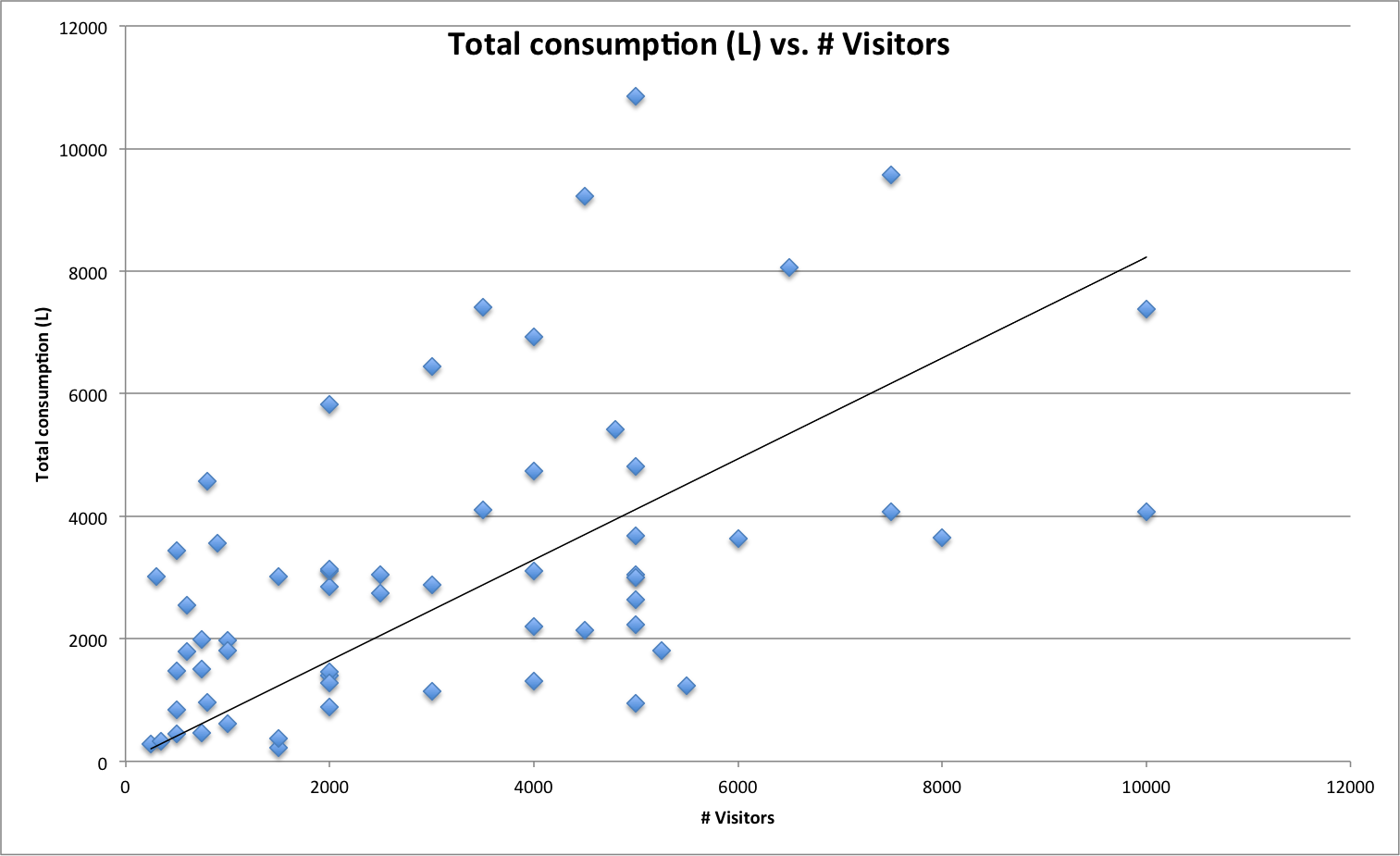
### 5.1.1 Expected number of visitors

The first factor that we will look at is the expected number of visitors. We will plot the total consumption in litres against the number of visitors to see what the influence on the total consumption is.



*Figure X: The total consumption (L) versus the number of visitors*

As shown in figure X, most of the data points are in the range of 0 to 10000 visitors, and 0 to 10000 litres consumed. There are some outliers with a lot of visitors and relatively little consumption and there are outliers with relatively few visitors but a lot of consumption. It might be interesting to take a closer look at the region where most of the data points are (figure X).

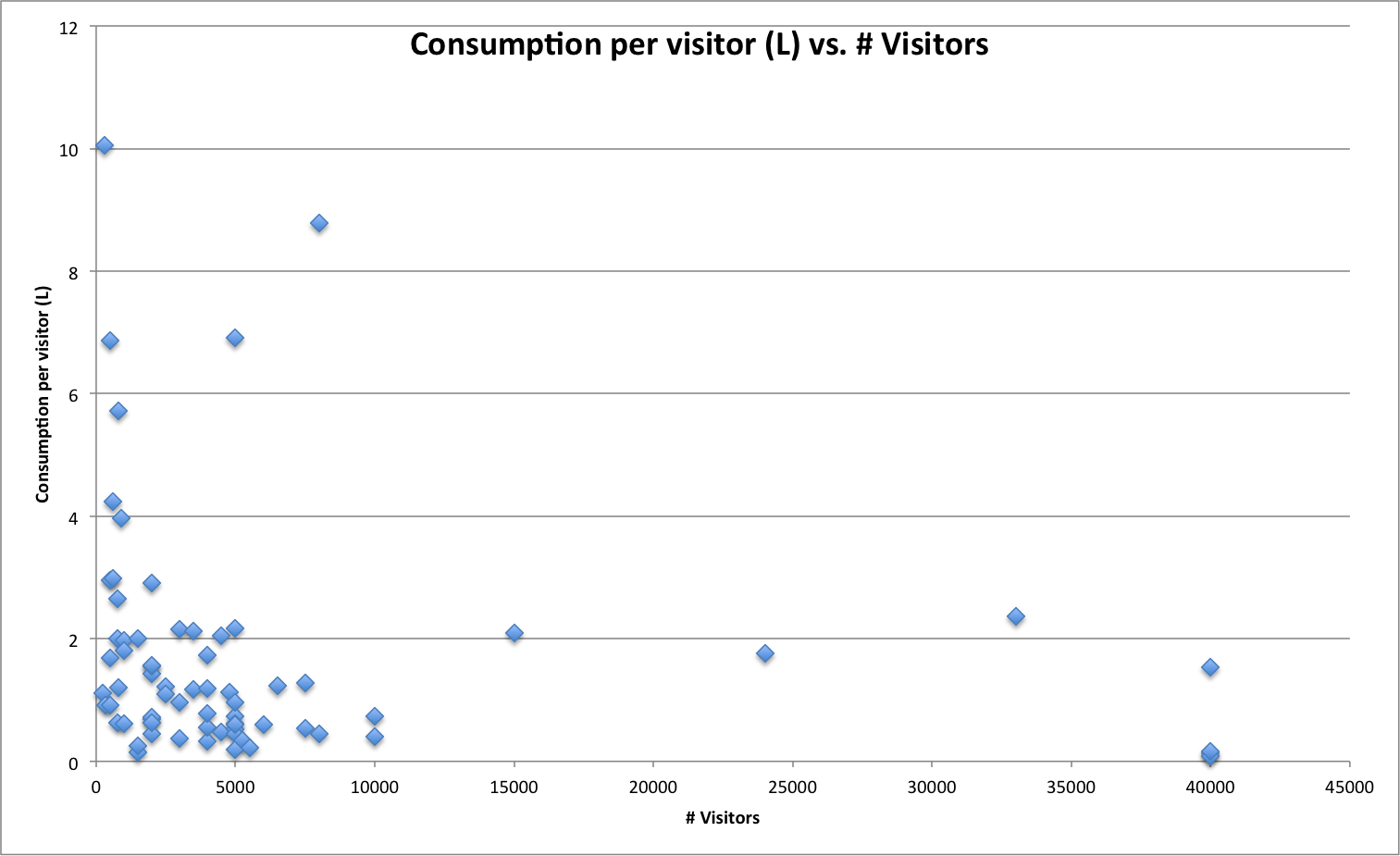


*Figure X: The total consumption (L) versus the number of visitors (zoomed in) with trend line*

There seems to be a linear trend, but the total consumption and the number of visitors do not seem to be very highly correlated. When we compute the correlation we get a value of 0,52. This confirms our expectations and tells us that the total consumption and the number of visitors are moderately correlated.

Something else that is interesting to look at is the average consumption per visitor against the number of visitors. That way we can determine if people consume more on average at bigger events, or at smaller events. The results are displayed in figure X.

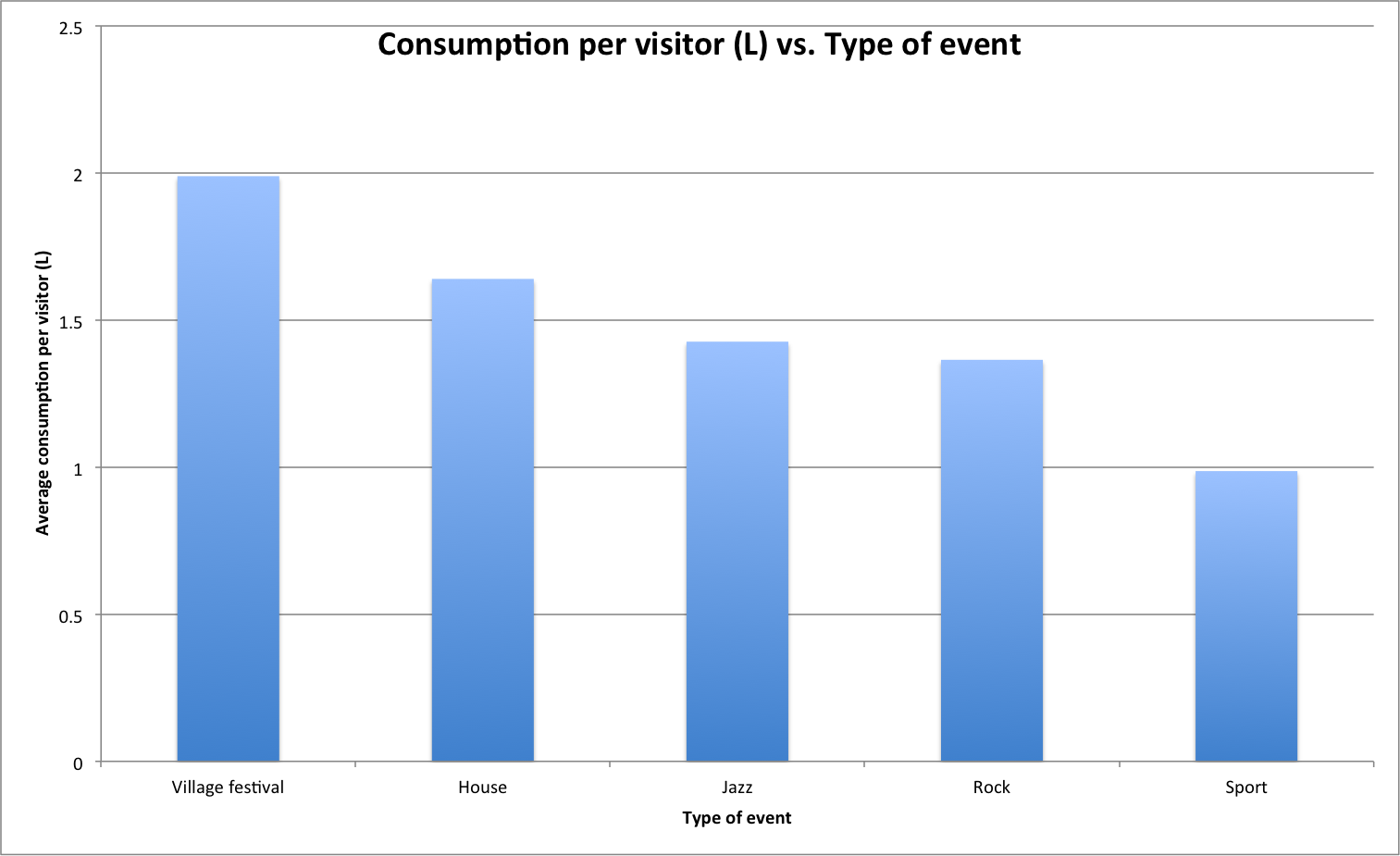
There does not seem to be a correlation between the number of visitors and the average consumption per visitor. For some smaller events there is a relatively high average consumption per visitor, with sometimes up to 10 litres consumed per visitor. This is not very likely and probably an error in the data. For large events the average consumption per visitor is not significantly higher than for small events. The correlation between the average consumption per visitor and the number of visitors is -0,16 and therefore tells us that there is no significant linear correlation.



*Figure X: Average consumption per visitor (L) versus the number of visitors*

### 5.1.2 Type of event

The next factor to be analysed is the type of event. To see if the type of event has a strong influence on the consumption, we will take a look at the average consumption per visitor for the different types of events.



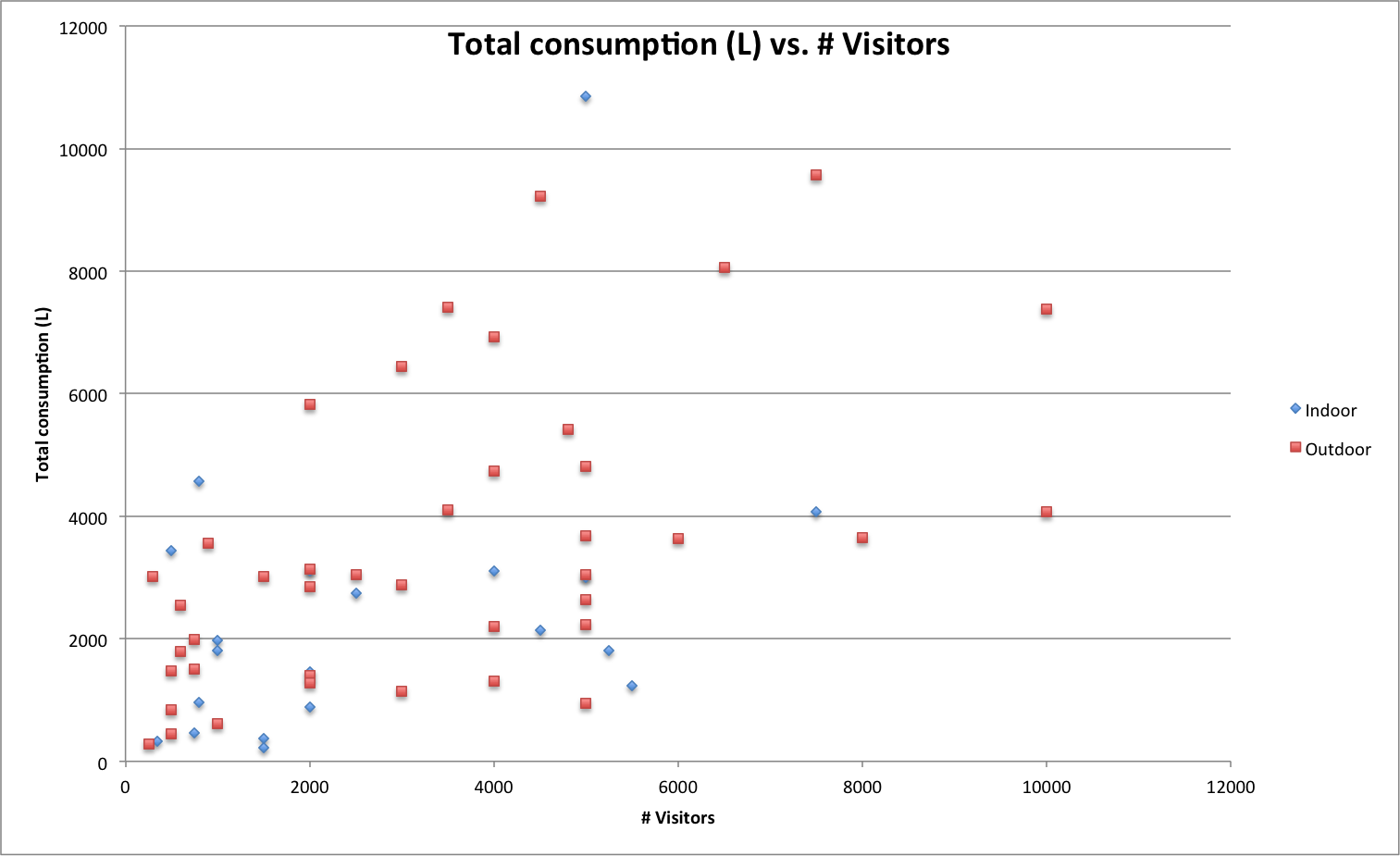
*Figure X: Average consumption per visitor (L) versus the type of event*

We can see in figure X that the average consumption at village festivals is significantly higher (approximately twice as high) than the consumption at sport events. At village festivals visitors are more likely to drink many beverages (especially beer), while at sport events there are a lot of people who bring their own drinks.

Between the music categories (house, jazz and rock) there is not a lot of difference in average consumption per visitor. A possible reason for the average consumption at house events being higher than at jazz and rock events is the fact that there is more consumption of water at house events due to people dancing and drug use. This will be broken down further in section X (…).

### 5.1.3 Indoor / Outdoor

Next we will investigate the influence of an event being indoors or outdoors on the average consumption. If the assumptions in chapter X are true, we expect to see a more distinctive linear relation between the number of visitors and the total consumption for indoor events, since there are fewer factors influencing the consumption.

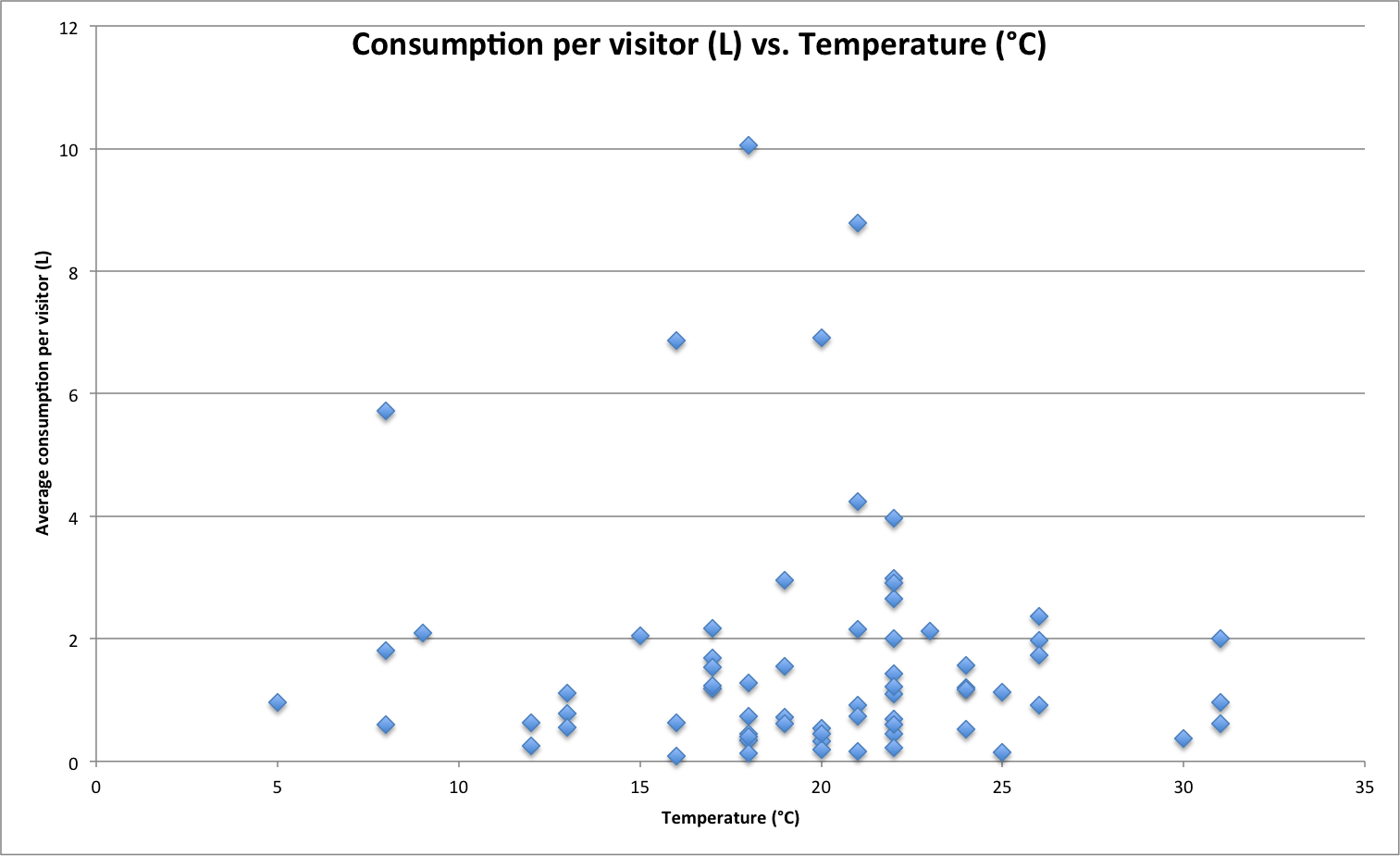


*Figure X: Total consumption (L) versus the number of visitors for indoor and outdoor events*

Figure X confirms that this assumption is true, because the data points of the outdoor events (the red squares) are more spread out than the data points of the indoor events (the blue diamonds). The correlation supports this conclusion: the correlation between the number of visitors and the total consumption for indoor events is 0,83, while it is only 0,50 for outdoor events. We can use this conclusion in a regression model.

### 5.1.4 Temperature

When the temperature is higher, it is possible that visitors of events will drink more because they get thirsty faster. Therefore we want to look at the influence of the temperature on the average consumption per visitor.

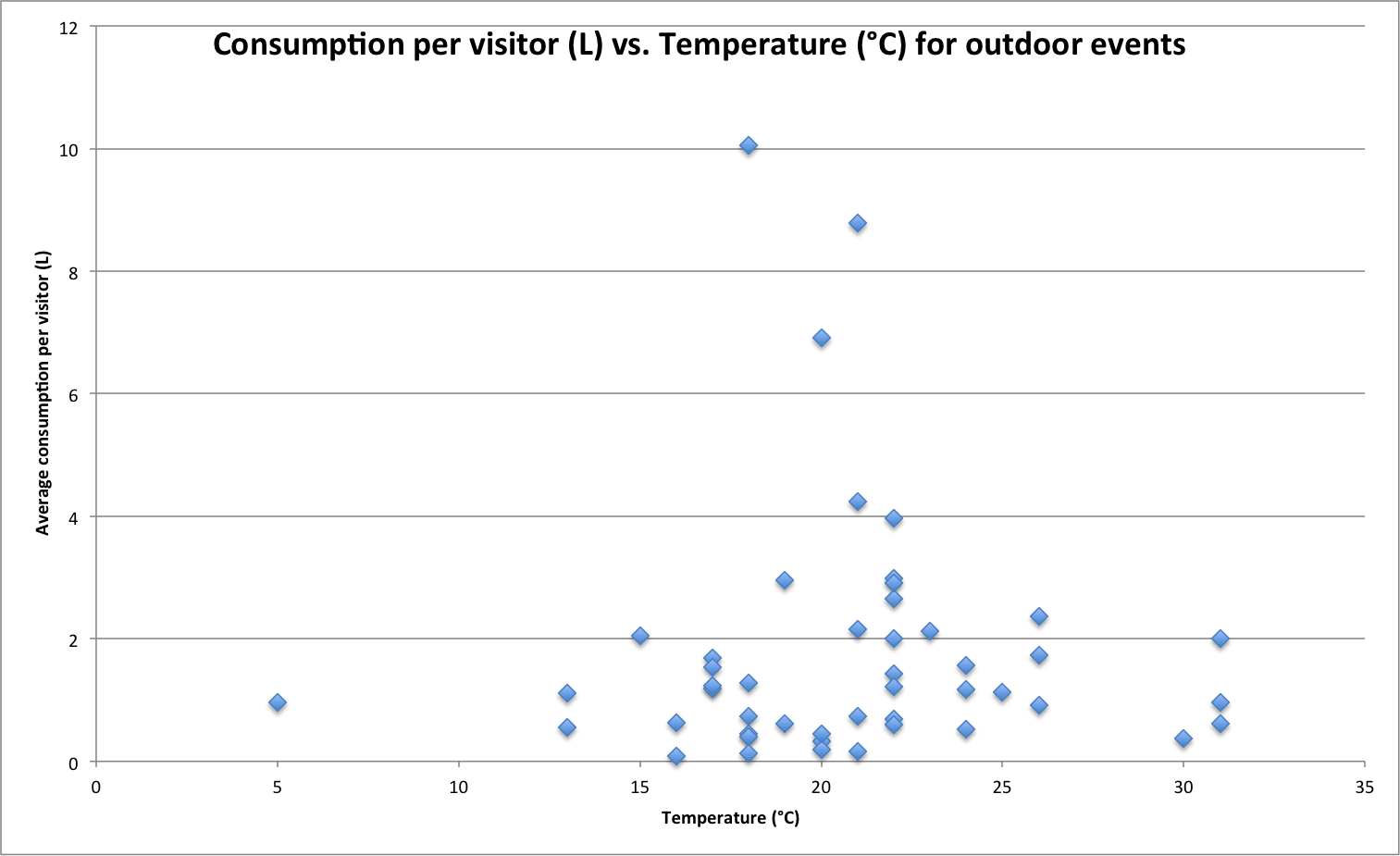


*Figure X: Average consumption per visitor (L) versus the temperature (°C)*

From the scatter plot it appears that there is no correlation at all between the temperature and the average consumption per visitor. The computed correlation turns out to be -0,06, so the influence of the temperature seems to be insignificant.

Since we have determined that an event being indoors or outdoors has an effect on the correlation between the number of visitors and the total consumption, it is interesting to investigate the influence of the temperature on just the outdoor events. There is a possibility that the indoor events are causing noise, so we will take a look at the average consumption per visitor against the temperature for the outdoor events only.

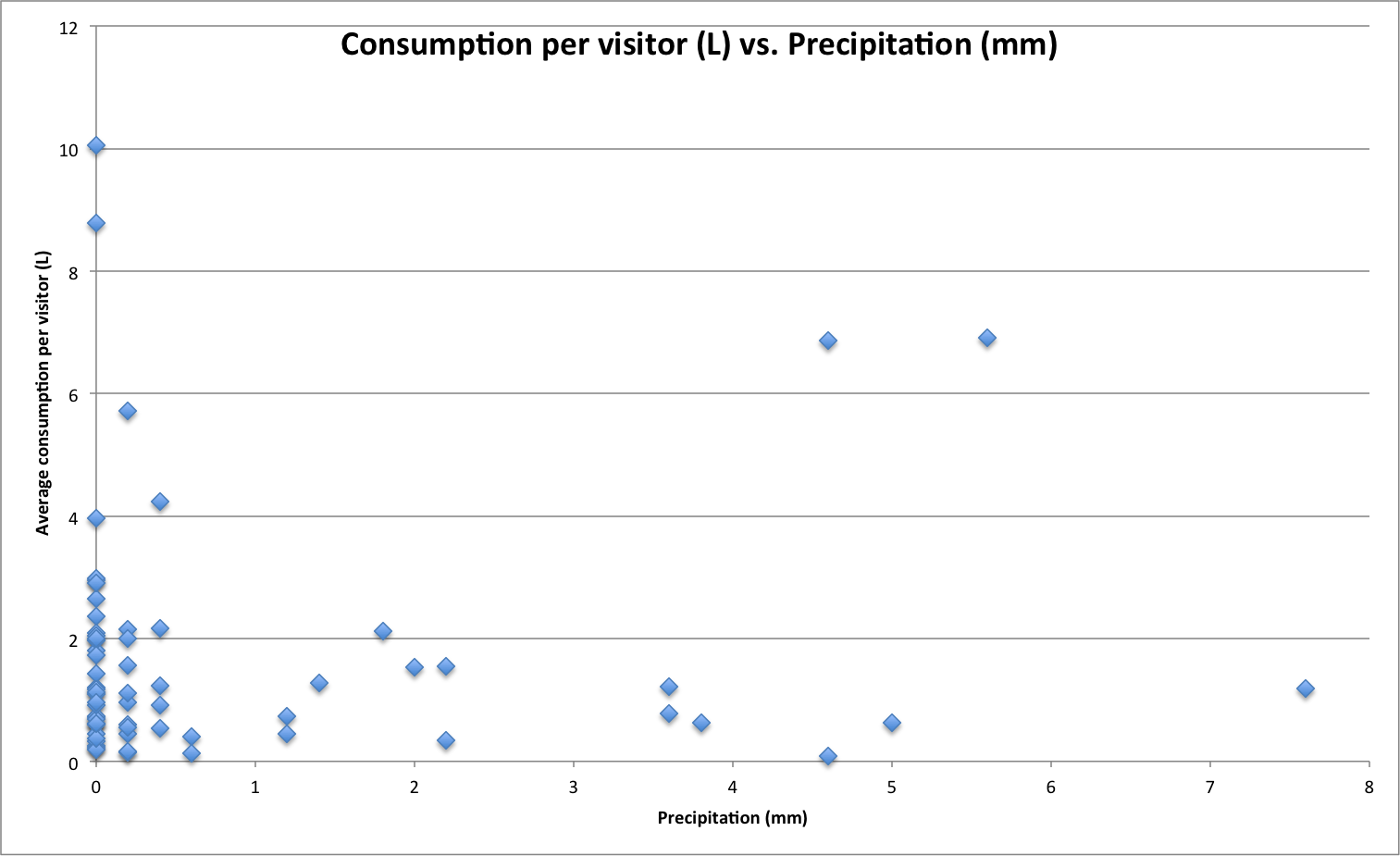
As shown in figure X, there is still no correlation between the temperature and the average consumption. The outliers are still present, and most of the other data points lie between 0 L per visitor and 2 L per visitor. The absence of correlation is confirmed when we compute the correlation, which is 0,003, indicating no correlation at all.



*Figure X: Average consumption per visitor (L) versus the temperature (°C) for outdoor events*

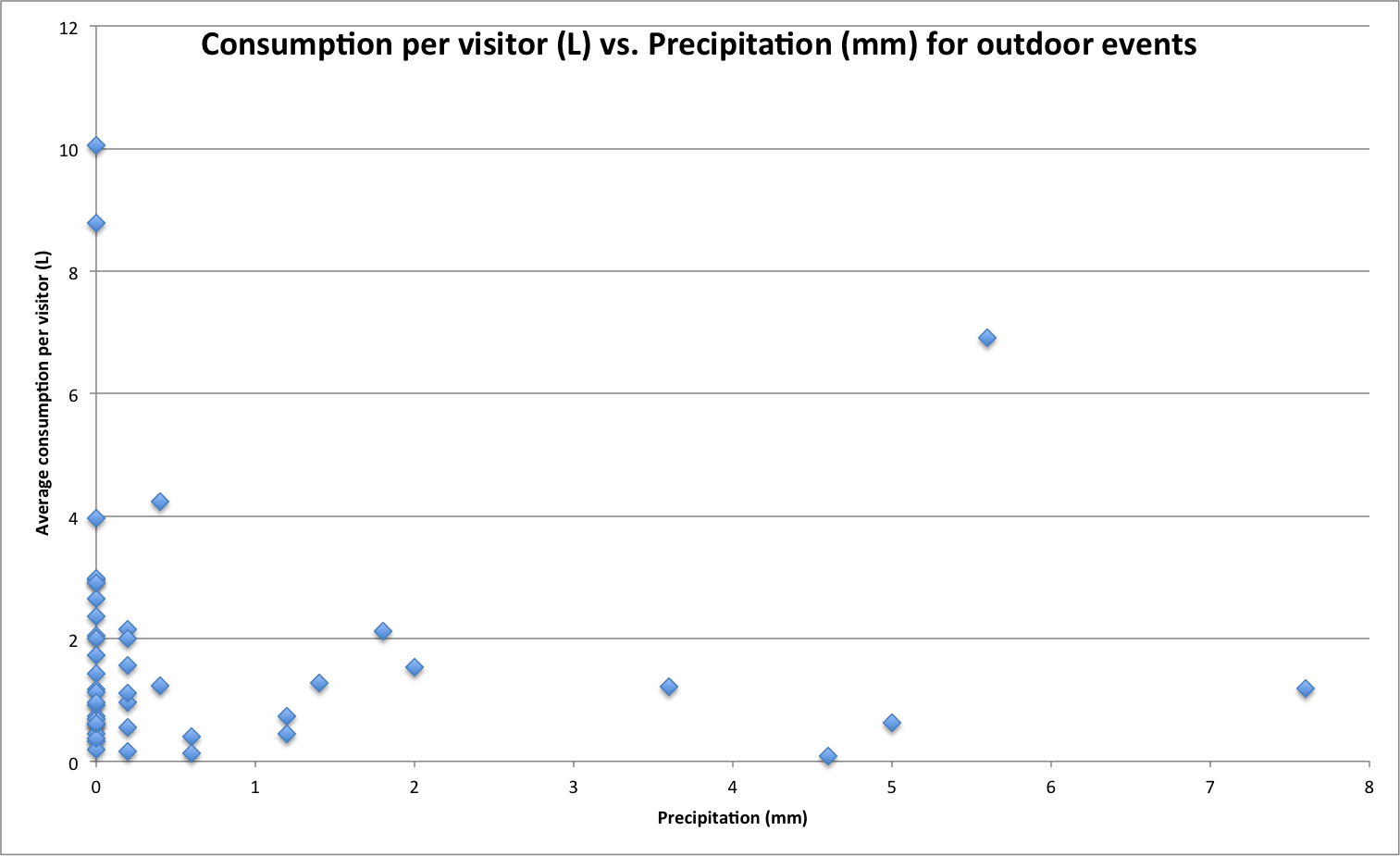
### 5.1.5 Precipitation

The last individual factor we will investigate, before we head on to combinations of factors and combinations of factors and categories, is the precipitation. We want to know if people consume more or less on days that it is raining than on days that it is not raining, or if it does not have any influence. Similar to the analysis on the temperature, the outdoor events are also analysed separately.



*Figure X: Average consumption per visitor (L) versus the precipitation (mm)*

What immediately stands out is the large number of events where the precipitation is 0 mm, meaning that there was no rain on that day. Because the spread for events with a precipitation of 0 mm is high, it is hard to discover correlation between the precipitation and the average consumption per visitor. Once again computing the correlation confirms this assumption; the correlation is only 0,08.



*Figure X: Average consumption per visitor (L) versus the precipitation (mm) for outdoor events*

In figure X the average consumption per visitor is plotted against the precipitation for the outdoor events. In the scatter plot we can see that not much has changed compared to the plot with the indoor events included. There are once again a lot of events without precipitation and the average consumption is spread between 0 L and 10 L per visitor. The correlation in this case is 0,014.

### 5.1.6 Conclusions

Based on these analyses we have drawn several conclusions that we can use for further analysis and building a prediction model. We will shortly summarize them conclusions:

* *The number of visitors and the total consumption are moderately correlated.*

When there are more visitors at an event, the total consumption is usually higher. There are events with a large number of visitors but a relatively low total consumption. A possible explanation for this is that the organisers have ordered a part of their order at another supplier. In that case the actual consumption was higher, but unfortunately we do not have insight into the data of other suppliers.

* *The type of event influences the average consumption per visitor.*

For certain types of events the average consumption is significantly higher than for other types of events. Therefore, when we want to predict the consumption for an event, we need to take the type of event into account.

* *For indoor events the correlation between the number of visitors and the total consumption is higher than for outdoor events.*

The correlation between the number of visitors and the total consumption is significantly higher for indoor events than for outdoor events. This might be caused by the circumstances for indoor events being more often similar to each other than the circumstances for outdoor events.

* *The temperature and precipitation do not influence the average consumption per visitor individually.*

We are not able to discover a clear linear relation between the average consumption per visitor and the temperature or the precipitation. Even when we disregard the indoor events (because the weather is less influential on indoor events) there is still a lack of correlation.

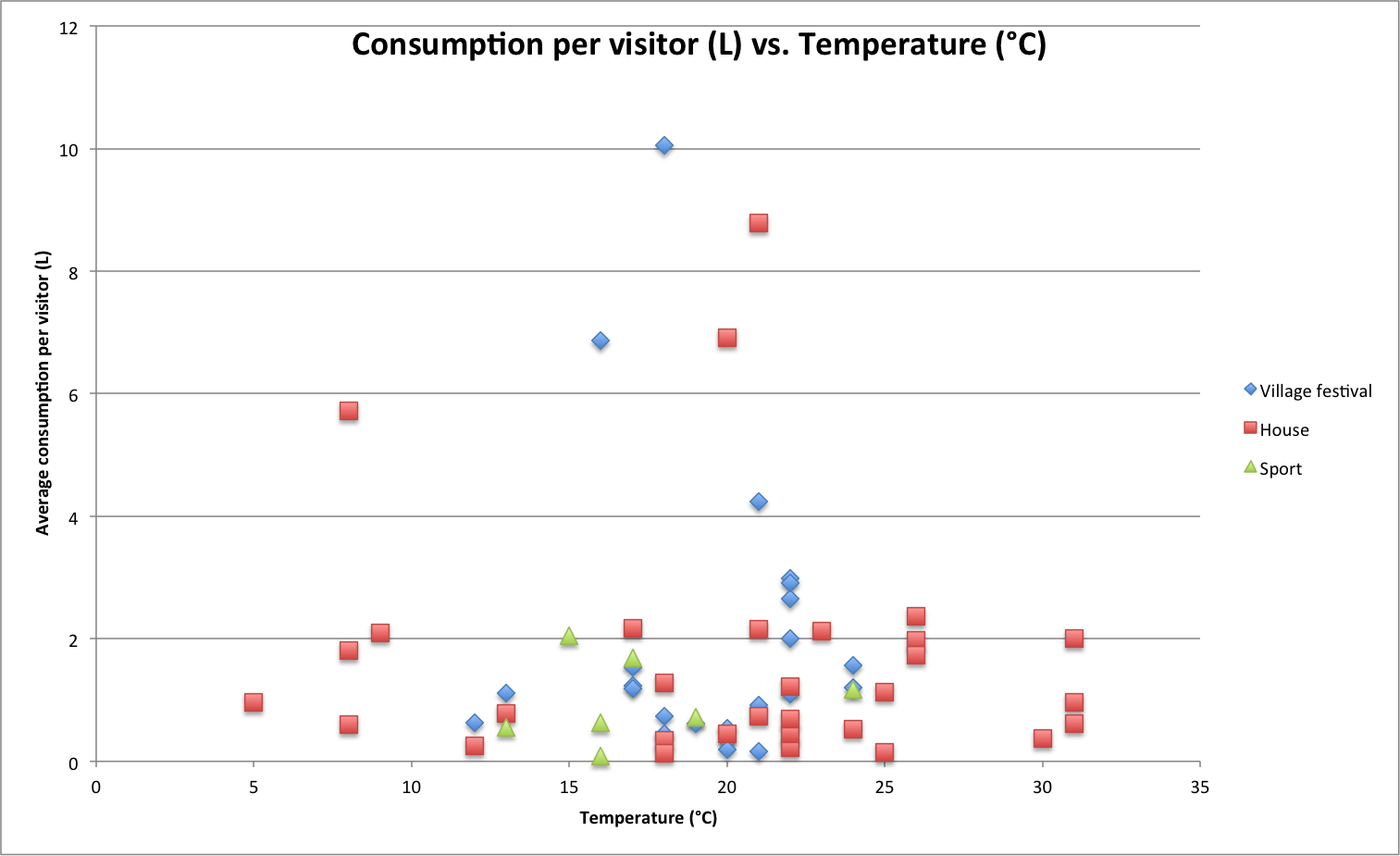
## 5.2 Influence of combinations

Now that we have determined the influence of individual factors, we can take a look at what happens if we combine factors. That way we can try to increase the usability of factors that did not have a significant influence when their performance was measured individually. For example, the temperature and the precipitation did not have a significant correlation with the average consumption, but combining them with the type of event might yield different results.

We are also going to investigate the influence of the individual factors on the consumption of the separate main categories. Different types of events attract a different audience, of which the preference for beverages can differ. So we want to investigate whether there is not only a higher average consumption at certain types of events, but if there is a significant difference in the consumption of the main categories too. The temperature and the precipitation do not have a correlation with the total consumption or the average consumption per visitor, but it is possible that they do influence the consumption of the main categories. When it is warm outside visitors might be inclined to drink other kinds of beverages than when it is cold.

### 5.2.1 Temperature and type of event

The first two factors we are combining are the type of event and the temperature. Although for some type of events the temperature will probably have no influence on the consumption, for other types it may play a role. Especially at sport events a higher temperature can cause an increase in the consumption of beverages. Since the consumption for jazz and rock events is similar to the consumption at house events and we do not have a lot of data available for different temperatures (the temperature did not differ a lot between jazz and rock events) we will take a look at the consumption at village festivals, house events and sport events.



*Figure X: Average consumption per visitor (L) versus the temperature (°C) for different types of events*

We notice that it looks like there has not changed much if we look at the average consumption for the village festivals (blue diamonds) and the house events (red squares). For the sport events (green triangles) there seems to be a small upward trend. When we compute the correlation, we get the following results:

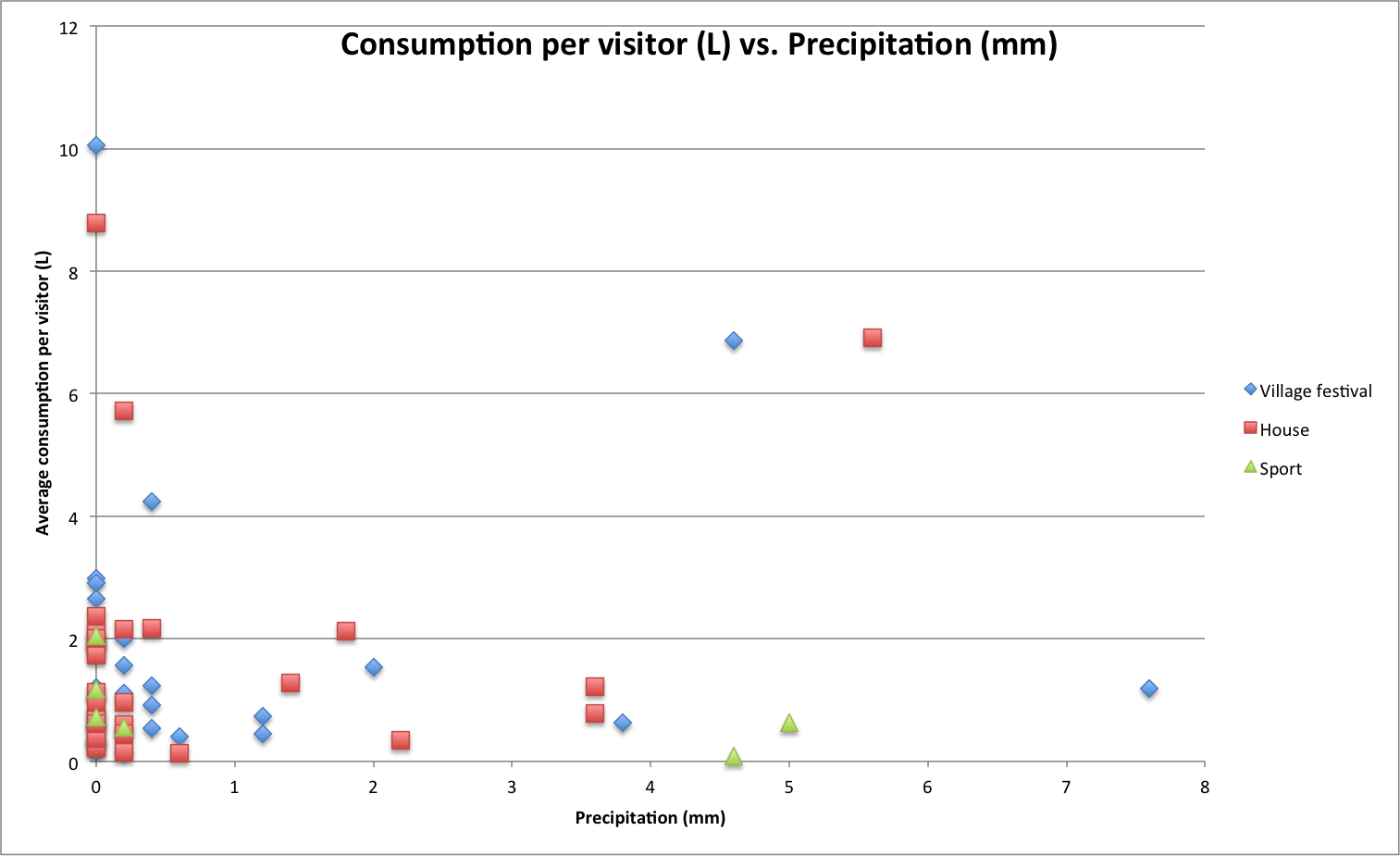
|  |  |
| --- | --- |
| Factor(s) | Correlation |
| Temperature | -0,06 |
| Temperature + Village festival | -0,04 |
| Temperature + House event | -0,11 |
| Temperature + Sport event | 0,11 |

*Table X: Correlation between average consumption per visitor and temperature for different types of events*

Combining the temperature with village festivals does not provide us with better results than when we consider only the temperature (the correlation becomes even smaller than it already was). When we look at a combination of the temperature and house events or sport events, there is an increase in correlation. Nevertheless the correlation is still negligible, even for a combination of the temperature and a type of event.

### 5.2.2 Precipitation and type of event

We can perform the same analysis on combinations of the precipitation and the type of event. Rain will probably not cause visitors to leave earlier at sport events (i.e. if someone bought a ticket for a football match, he / she will probably not leave before the match ends, regardless of rainfall). At village festivals, however, people may be inclined to go home earlier if the weather is bad.



*Figure X: Average consumption per visitor (L) versus the precipitation (°C) for different types of events*

Similar to what we saw in the analysis of the individual factors, a lot of the data points are clustered around a precipitation of 0 and 0,2. For some events the spread is relatively small, which can indicate a higher correlation between the average consumption per visitor and the precipitation.

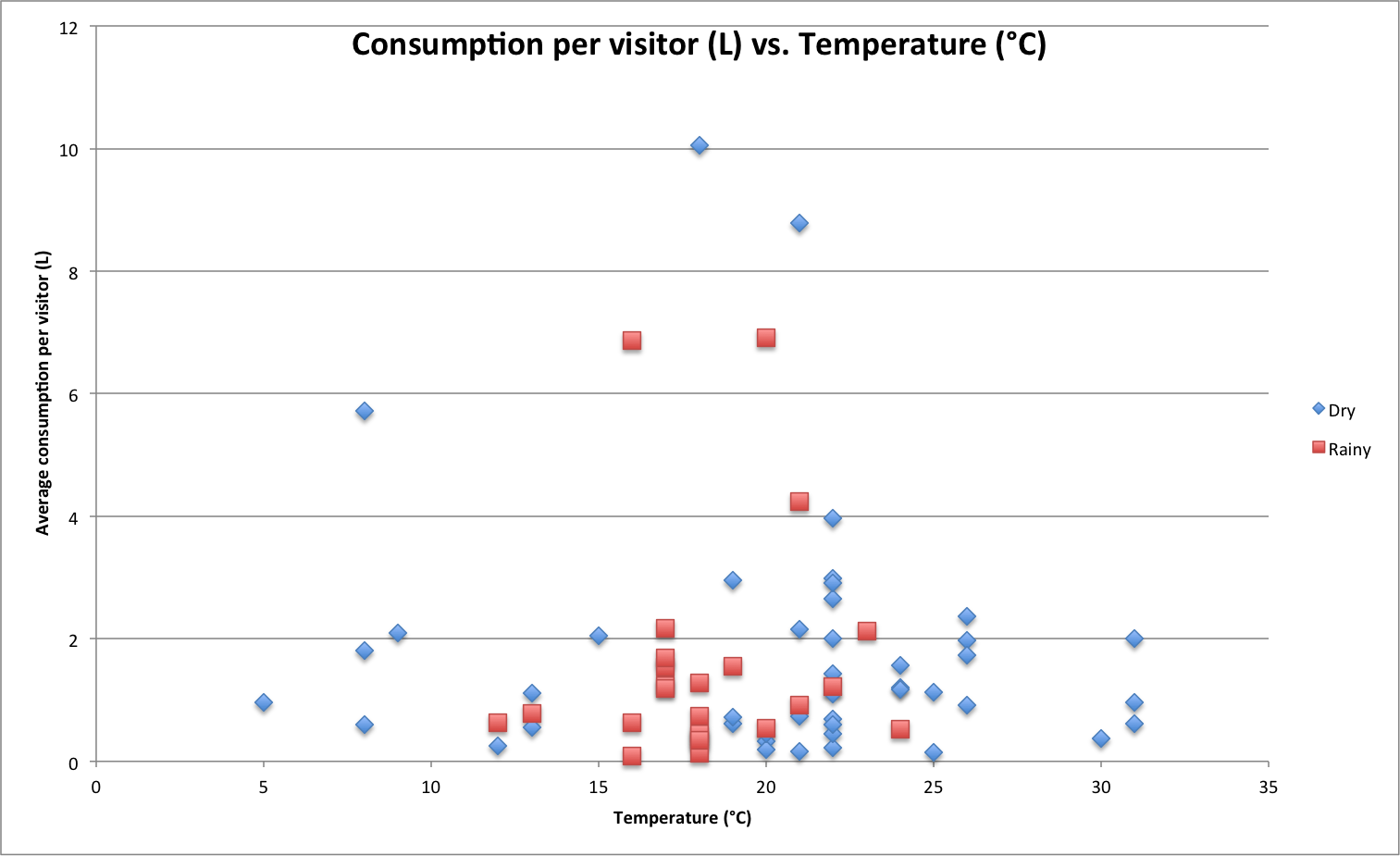
|  |  |
| --- | --- |
| Factor(s) | Correlation |
| Precipitation | 0,08 |
| Precipitation + Village festival | 0,07 |
| Precipitation + House event | 0,26 |
| Precipitation + Sport event | 0,20 |

*Table X: Correlation between average consumption per visitor and precipitation for different types of events*

These are interesting results. Where we would expect to see a negative correlation for the combination of precipitation and the village festivals, but instead we see an increase in positive correlation when we combine the precipitation with the house events and the sport events. This might mean that instead of people leaving earlier at village festivals due to bad weather, people stay longer at house and sport events because they do not want to leave yet. Although it has to be mentioned that the correlation is still relatively low, so we should be careful to draw conclusions based on these results.

### 5.2.3 Temperature and precipitation

To combine the temperature and the precipitation, we classify the precipitation. We assign every event to on of two classes; we assign the event to the class ‘Dry’ if the precipitation was 0,2 mm or less, or to the class ‘Rainy’ if the precipitation was more than 0,2 mm. Then we perform our analysis.



*Figure X: Average consumption per visitor (L) versus the temperature (°C) for different levels of precipitation*

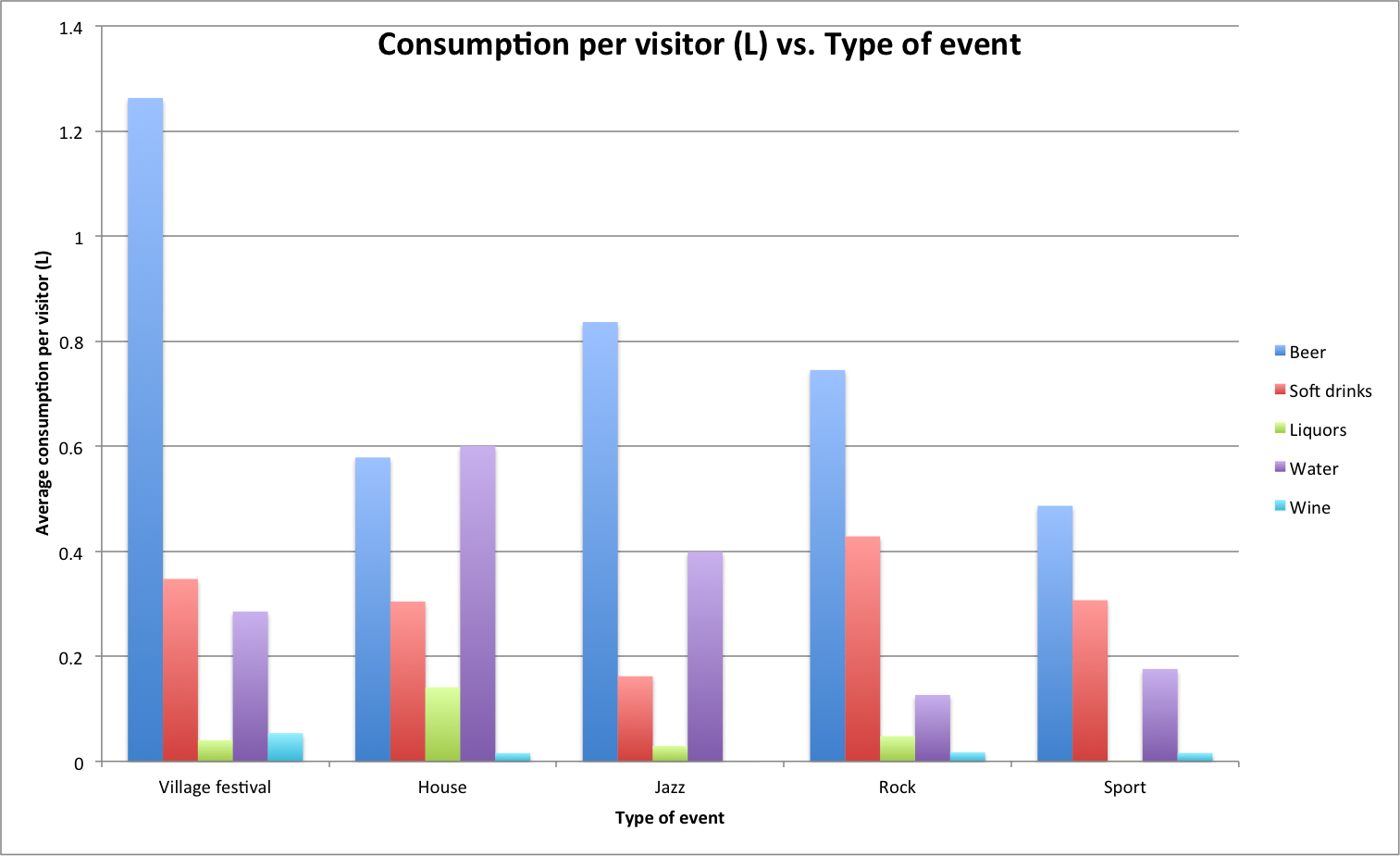
For both precipitation classes there does not seem to be a significant correlation. For events with dry weather a small downward trend can be detected and for events with rainy weather there is a small upward trend, but this trend is negligible. This can also be seen in table X, with the correlations.

|  |  |
| --- | --- |
| Factor(s) | Correlation |
| Temperature | -0,06 |
| Temperature + Dry weather | -0,11 |
| Temperature + Rainy weather | 0,11 |

*Table X: Correlation between average consumption per visitor and temperature for different types of weather*

### 5.2.4 Type of event and main categories

We have already observed that the type of event influences the average consumption per visitor. We suspect that there might also be a relation between the type of event and the consumption of the different main categories. For each type of event we will look at the average consumption of the main categories.

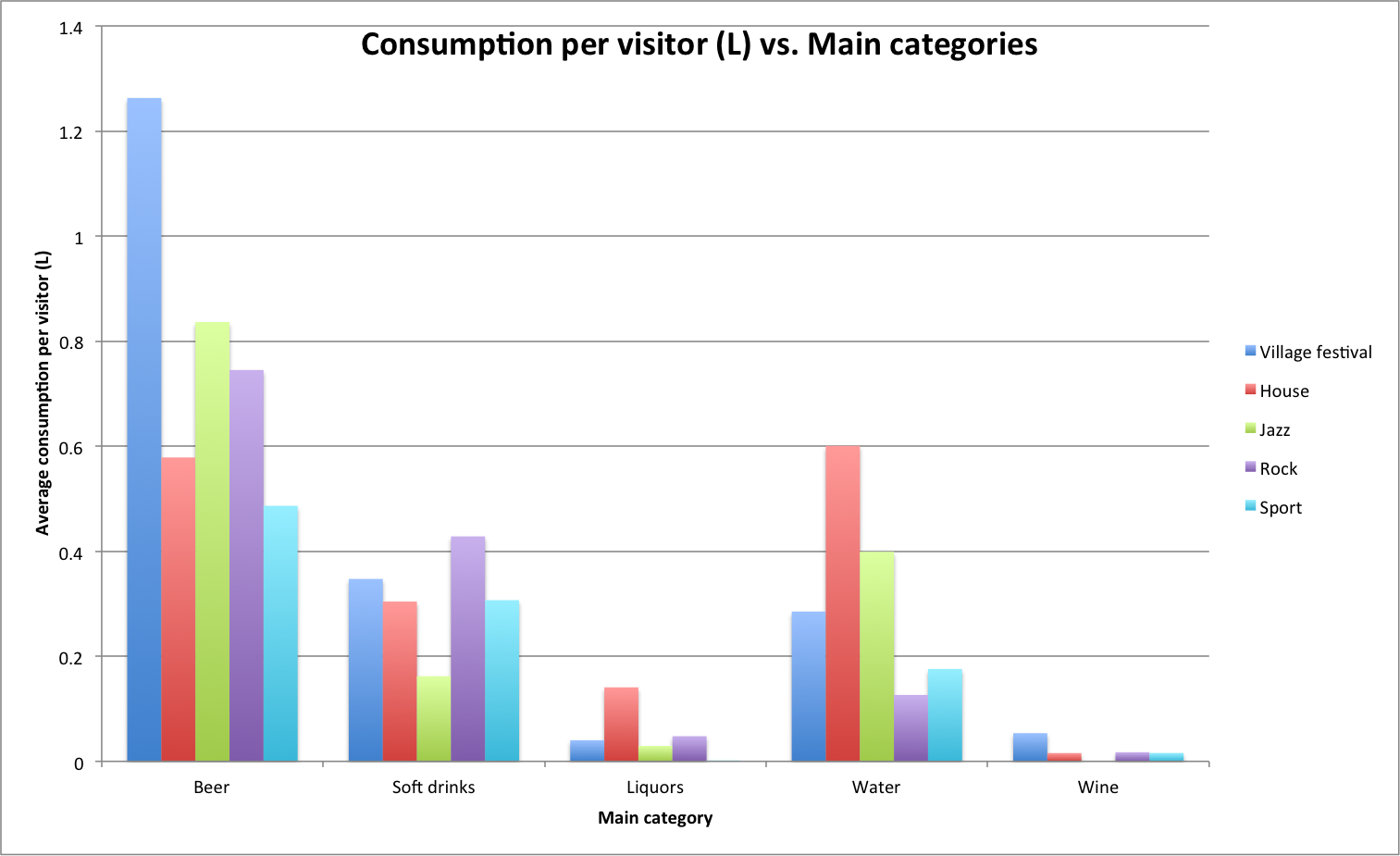


*Figure X: Average consumption per visitor (L) versus the type of event for the main categories*

We can see in figure X that for the village festivals there is one clear category that stands out. The beer consumption is more than three times as high as the consumption of soft drinks and water. For most events the beer consumption is higher than the consumption of other categories.

Only for the house events the average water consumption is higher than the beer consumption. Since people are more often dancing at house events they dehydrate faster, so they will drink more water than visitors of other types of events. Drug use may also play a role in the relatively high average consumption of water.

For rock events and sport events the consumption of soft drinks is the second highest category in terms of average consumption. Since energy drinks are a subcategory of soft drinks, the average consumption of soft drinks is expected to be relatively high. There will also be more children attending sport events than for example house events, which also explains the fact that the consumption of soft drinks at sport events is higher. For rock events there is not such an obvious explanation, although it is possible that there are more children attending rock events. Most house events have an age limit while this is not the case for most rock events.



*Figure X: Average consumption per visitor (L) versus the main categories for the types of events*

When looking at the average consumption per visitor for the different main categories there are two main categories that stand out. For almost every type of event the consumption of liquors and wine is very low.

Because the average consumption is in litres, the amount of liquors consumed is much lower than the average consumption of categories as beer and soft drinks. Liquors are often mixed with soft drinks at house events, explaining the higher average liquor consumption at house events relative to the other types of events.

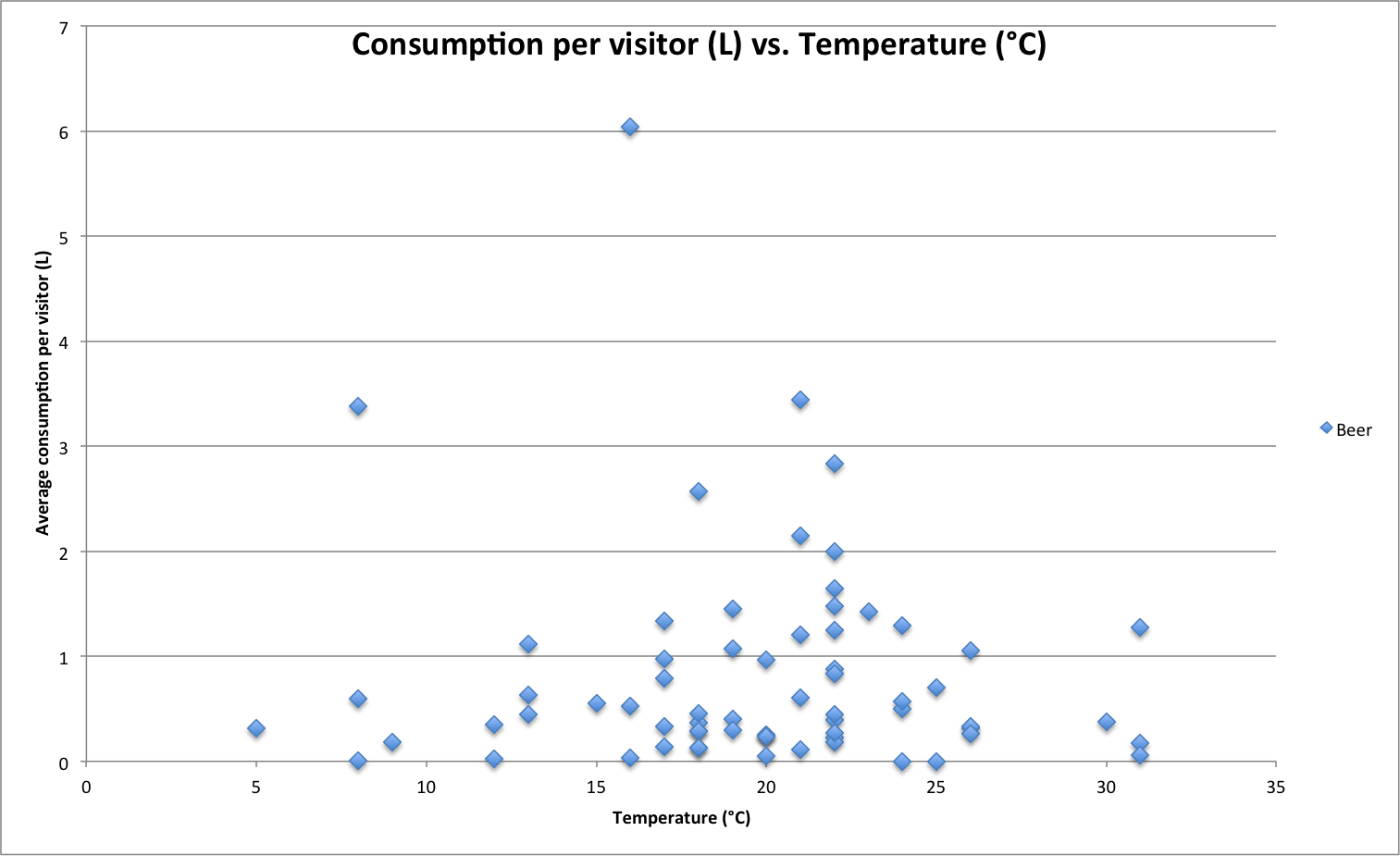
Since there are a lot of wine suppliers (e.g. so-called ‘wine houses’), the average consumption of wine is low. Village festivals are often organised by café owners, who sometimes decide to order the wines at Heineken, but most of the times organisers decide to order the wines for their event somewhere else.

### 5.2.5 Temperature and main categories

Finally, we will investigate the influence of the temperature on the consumption of the different main consumption categories. The temperature might not be correlated with the average consumption for all of the main categories together, but it is possible that it is correlated with some of the separate main categories.

Because the average consumption of some main categories is a lot higher than the average consumption of other categories, plotting the average consumption of each separate category against the temperature will give us a plot in which the data for the liquor and wine consumption will be barely visible. Therefore we make three separate plots: one with the average beer consumption versus the temperature, one with the soft drinks and water consumption, and one with the liquor and wine consumption.

First we will take a look at the average beer consumption per visitor for different temperatures.

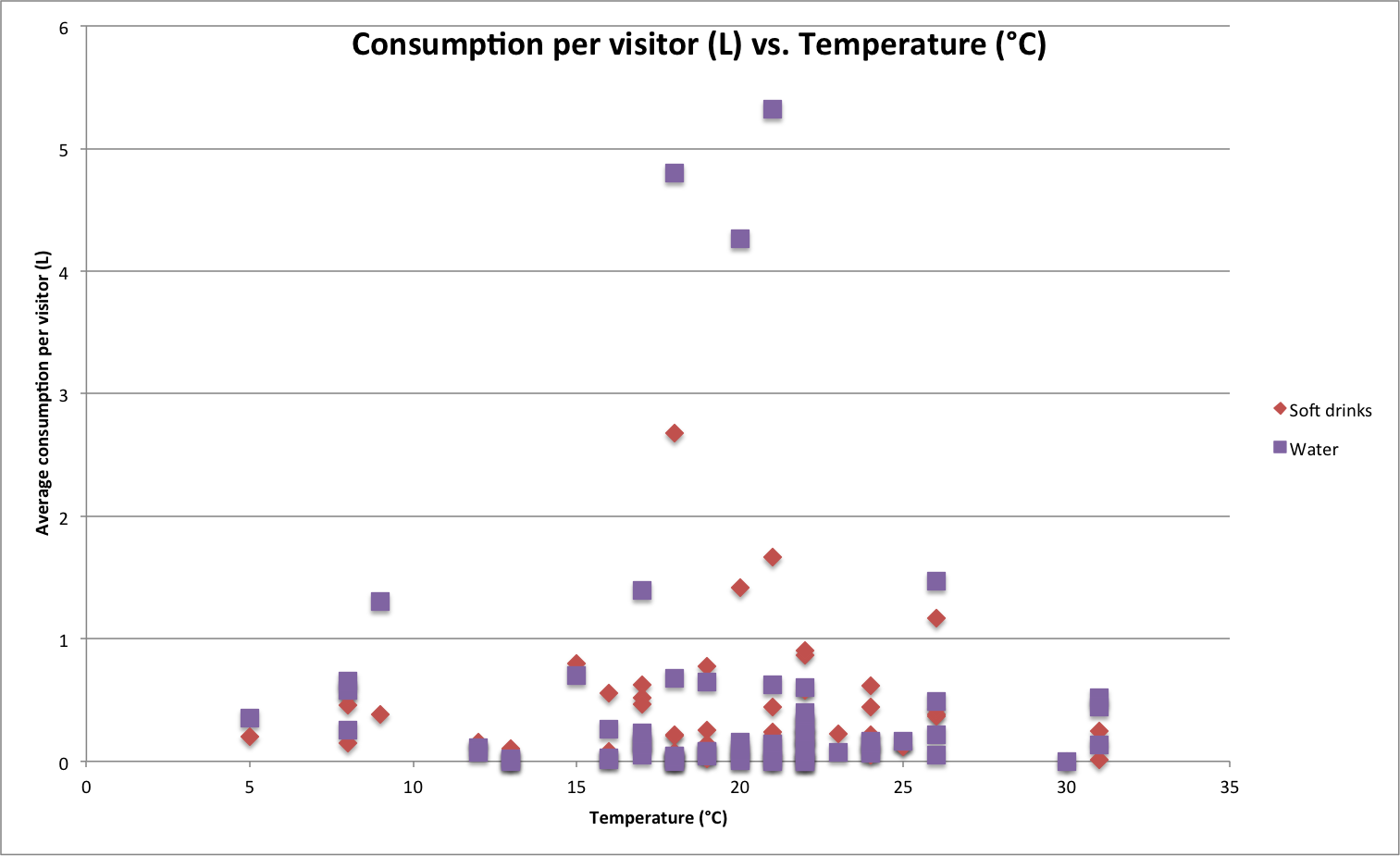


*Figure X: Average beer consumption per visitor (L) versus the temperature (°C)*

When we compare this scatter plot to figure X, the average consumption per visitor for all the main categories together versus the temperature, we notice that they look very much alike. This indicates that it is unlikely that there is correlation between the temperature and the average beer consumption per visitor.

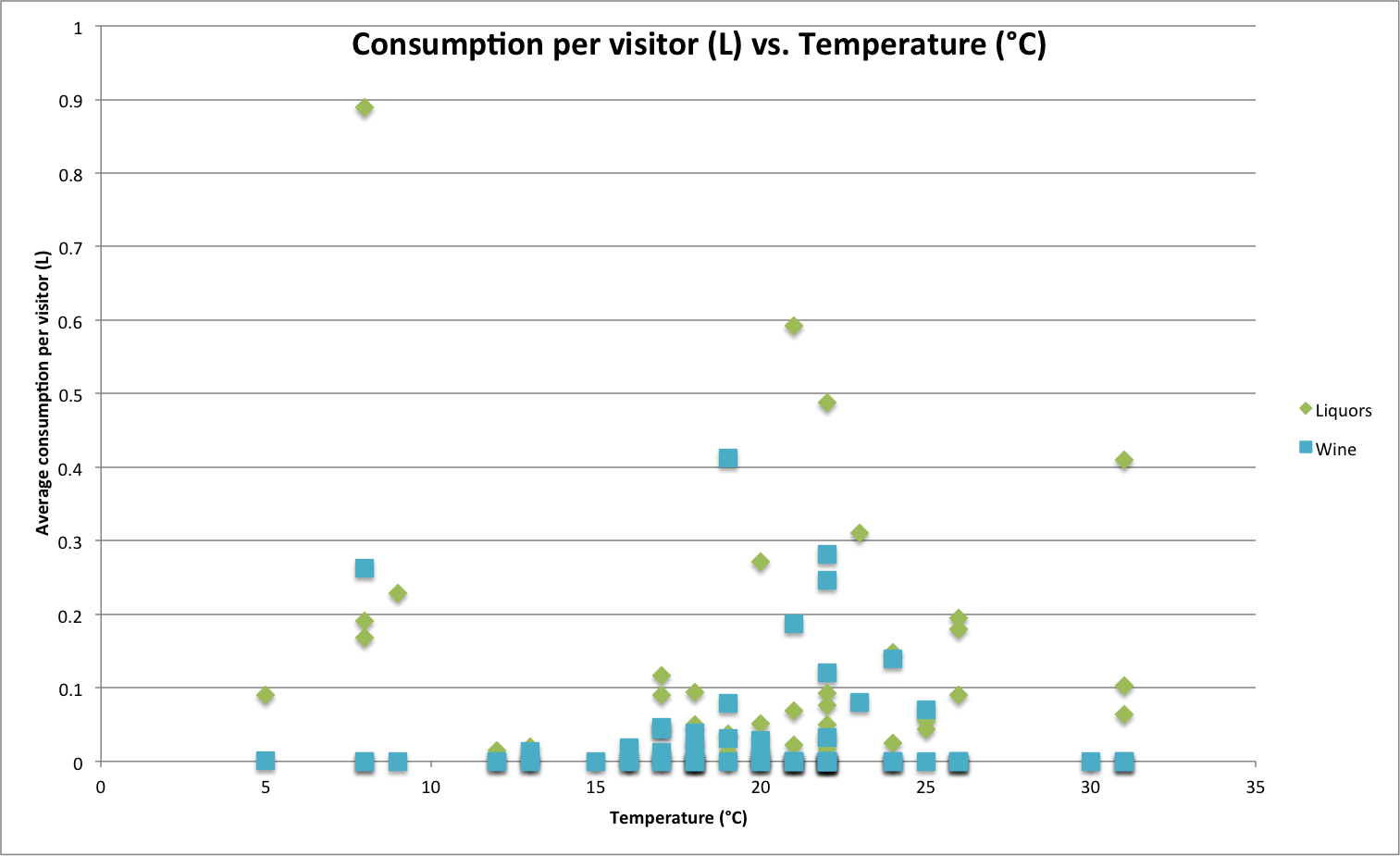
Next we look at the average consumption of soft drinks and water. Perhaps the visitors will not consume more beer when it is warmer, but do they choose do drink more water instead.

In figure X, the red diamonds represent the average soft drink consumption per visitor and the purple squares represent the average water consumption per visitor. This chart does not confirm our expectation that visitors consume more water when it is warmer than when it is cold; except for a few outliers the average consumption is approximately the same for every temperature. The same goes for soft drinks, for which the average consumption is also approximately the same for the different temperatures.



*Figure X: Average soft drink and water consumption per visitor (L) versus the temperature (°C)*

The last two categories we will look at are the liquors and the wines. The average consumption per visitor of these categories is shown in figure X. The average liquor consumption is displayed with green diamonds and the average wine consumption is displayed with light blue squares.



*Figure X: Average liquor and wine consumption per visitor (L) versus the temperature (°C)*

For the liquors and the wines the same pattern appears as seen in the charts for the average beer and the soft drink and water consumption. There are once again a few outliers, but there is no clear upward or downward trend.

The computed correlations (displayed in table X) confirm the conclusions we have drawn from the scatter plots:

|  |  |
| --- | --- |
| Factor(s) | Correlation |
| Temperature Overall | -0,06 |
| Temperature + Beer | -0,07 |
| Temperature + Soft drinks | -0,004 |
| Temperature + Liquors | -0,09 |
| Temperature + Water | -0,03 |
| Temperature + Wine | -0,03 |

*Table X: Correlation between average consumption per visitor and temperature for the different main categories*

It is clear that the temperature is not correlated to any of the main consumption categories. For higher temperatures the average consumption per visitor for each main category is approximately the same as for lower temperatures, and for each category there are some outliers.

### 5.2.6 Conclusions

After analysing several combinations of factors and combining factors with the main categories, we can draw several interesting conclusions:

* *The weather has a negligible influence on the average consumption of visitors*

Combining the weather factors (temperature and precipitation) with the type of event of with each other increases the correlation with the average consumption per visitor in most cases, but the correlation is not significant. The correlation is never higher than 0,3, indicating that the weather does not play a significant role in the consumption behaviour of visitors of the analysed events.

* *Visitor consumption patterns are (very) different for the different types of events*

For different types of events, there are large differences in the consumption of the main categories. Rock events and sport events show a similar consumption pattern, but all of the other event types have a very different consumption pattern. In the prediction model we need to take these differences into account.

## 5.3 Trends

There is still a lot of innovation in the field of alcoholic and non-alcoholic beverages. Until a few years ago most people in the Netherlands had never heard of Radler, but now it is almost everywhere available. Specialty beers are also growing in popularity. The increase in popularity of these new product categories is of course at the expense of existing product categories. Since this is something we might want to account for in our prediction model we will investigate the presence of trend in our data.

Because we only have data for one year of consumption, it will be hard to say anything about long-term trend. However, we can take a look at seasonal trends in our data. Some subcategories, e.g. Radler and beach beers are thought of as summer beers, while specialty beers are more often promoted in the winter. In the next sections the presence of these seasonal trends will be investigated.

### 5.3.1 Trends in main categories

B

### 5.3.2 Trends in subcategories

C

### 5.3.3 Conclusions

D

# 6. Building the tool

Text

# 7. Results

Text

# 8. On-Time project

Text

# 9. Conclusions and further research

Text

# Appendix

Text