# Case_Study_Group_24

The fictitious company “106” sells gearshift systems to car manufacturers. In order to improve the product of the automatic gearshift system “K3AG1”, the company plans to analyze the damage cases of the last years. The automatic gearshift “K3AG1” is a standard component of the car brand “OEM1” and can be selected by the customer as an equipment feature of the car types “Typ11” or “Typ12”.

This project serves to analyze the damage cases of the last years for the product improvement of the automatic gearshift system.

# 1. Data Import and Data Preparation

To reduce the import time, tidying is already performed according to the relevant columns of the respective data sets.

## 1.1 Relevant files

From all the files provided, the following files are imported for this case study:

- Fahrzeug:
  - Fahrzeuge_OEM1_Typ11.csv
  - Fahrzeuge_OEM1_Typ12.csv
  - Bestandteile_Fahrzeuge_OEM1_Typ11.csv
  - Bestandteile_Fahrzeuge_OEM1_Typ12.csv
- Komponente:
  - Komponente_K3AG1.csv
- Einzelteil:
  - Einzelteil_T21.csv
  - Einzelteil_T24.txt
  - Einzelteil_T25.csv
- Zulassungen:
  - Zulassungen_alle_Fahrzeuge.csv
- Geodaten:
  - Geodaten_Gemeinden_v1.2_2017-08-22_TrR.csv

## 1.2 Libraries

The following libraries are relevant for the import and data analysis and must therefore be loaded first.

```
if(!require("install.load")){
   install.packages("install.load")
}
library('install.load')
install_load("readr", "readxl", "dplyr", "stringr", "tidyr", "tidygeocoder")
```

# 1.3 Fahrzeug Typ 11 and Typ 12

The files are imported with read_csv and read_csv2 and the distinction comes from the different separators. Then the two tables are linked, which is possible because of the matching column names. The column “Fehlerhaft” has been renamed to “Fehlerhaft_Fahrzeug” to be able to distinguish later between the different “Fehlerhaft” columns. This allows us to clearly say whether the error is in the vehicle or in the component or individual parts. In addition, a new column is created to indicate the vehicle type. The vehicle type is part of the ID number. Finally, only the relevant columns are selected for data analysis.

```
Typ11 <- read_csv("Data/Fahrzeug/Fahrzeuge_OEM1_Typ11.csv")
Typ12 <- read_csv2("Data/Fahrzeug/Fahrzeuge_OEM1_Typ12.csv")

Fahrzeuge <- bind_rows(Typ11, Typ12) %>%
  rename("Fehlerhaft_Fahrzeug" = Fehlerhaft) %>%
  mutate(Fahrzeug_Typ = substr(ID_Fahrzeug, 1, 2)) %>%
  dplyr::select("ID_Fahrzeug", "Fehlerhaft_Fahrzeug", "Fahrzeug_Typ")
```

# 1.4 Bestandteile Fahrzeug Typ 11 and Typ 12

The files are imported and linked to each other. The table contains different gears, but for our analysis only the gear K3AG1 is relevant and accordingly only the rows containing this gear number are selected. Then the relevant columns were selected.

```
Teile_Typ11 <- read_csv2("Data/Fahrzeug/Bestandteile_Fahrzeuge_OEM1_Typ11.csv")
Teile_Typ12 <- read_csv2("Data/Fahrzeug/Bestandteile_Fahrzeuge_OEM1_Typ12.csv")

Teile_Fahrzeuge <- bind_rows(Teile_Typ11, Teile_Typ12) %>%
  filter(str_detect(ID_Schaltung, "K3AG1")) %>%
  dplyr::select(ID_Schaltung, ID_Fahrzeug)
```

# 1.5 Bestandteile Komponenten K3AG1

The given table is result of joining tables with bind_cols. Therefore columns occur several times and are distinguished e.g. with .x, but thereby many NA values exist. Therefore the table is divided into three subsets and all subsets get the same column names. Afterwards the subsets can be linked together with bind_rows and the NA values can be removed. Then the relevant columns were selected.

```
Komp  <- read_csv("Data/Komponente/Komponente_K3AG1.csv")

Komp_p1 <- subset(Komp, select = c(3:9))
Komp_p2 <- subset(Komp, select = c(10:16))
Komp_p3 <- subset(Komp, select = c(17:23))

col_names <- c("ID_Schaltung", "Produktionsdatum_Komp", "Herstellernummer_Komp", "Werks
nummer_Komp", "Fehlerhaft_Komp", "Fehlerhaft_Datum_Komp", "Fehlerhaft_Fahrleistung_Kom
p")

Komp_list <- list(Komp_p1, Komp_p2, Komp_p3)
Komp_list <- lapply(Komp_list, function(Komp) {
  colnames(Komp) <- col_names
  Komp
})

Komp_K3AG1 <- bind_rows(Komp_list) %>%
  filter_all(any_vars(!is.na(.))) %>%
  dplyr::select("ID_Schaltung", "Fehlerhaft_Komp")
```

# 1.6 Bestandteile Komponenten K3AG1

The file is imported and ID_K3AG1 is renamed so that it can be joined to another table later as the name of the key must be identical. Then the relevant columns were selected.

```
Teile_Komp_K3AG1  <- read_csv2("Data/Komponente/Bestandteile_Komponente_K3AG1.csv") %>%
  rename("ID_Schaltung" = ID_K3AG1) %>%
  dplyr::select(c(2:5))
```

# 1.7 Einzelteil T21

The file is imported and “Fehlerhaft” is renamed so it is clear that this column belongs to “Einzelteil T21”. Then the relevant columns were selected.

```
T21 <- read_csv2("Data/Einzelteil/Einzelteil_T21.csv") %>%
  rename("Fehlerhaft_T21" = Fehlerhaft) %>%
  dplyr::select("ID_T21", "Fehlerhaft_T21")
```

# 1.8 Einzelteil T24

The file is imported with read_file, which reads a complete file into a single object. In our string we can see that the line break is done with Form Feed, which will be replaced with Carriage Return and Line Feed. Then we read our string with read_delim. The values in the lines are separated with two spaces. The file has column headers, but only 22 and all other lines contain 23 columns. Accordingly, we have to skip the first line and adjust the column names. For this, the column name “Nummer” is added at the front and all other column names are kept. Then the table must be divided into three subsets, because columns occur several times and were not linked correctly. The subsets get the same column names and get linked with bind_rows and NAs are removed. Then the relevant columns were selected.

```
T24_string <- read_file("Data/Einzelteil/Einzelteil_T24.txt")
T24_string <- str_replace_all(T24_string, "\f", "\r\n")

T24 <- read_delim(T24_string, delim = "  ", skip = 1, col_names = c("Nummer", paste0
("V", 1:22)))

T24_p1 <- subset(T24, select = c(3:9))
T24_p2 <- subset(T24, select = c(10:16))
T24_p3 <- subset(T24, select = c(17:23))

col_names <- c("ID_T24", "Produktionsdatum_T24", "Herstellernummer_T24", "Werksnummer_T
24", "Fehlerhaft_T24", "Fehlerhaft_Datum_T24", "Fehlerhaft_Fahrleistung_T24")

T24_list <- list(T24_p1, T24_p2, T24_p3)
T24_list <- lapply(T24_list, function(T24) {
  colnames(T24) <- col_names
  T24
})

T24 <- bind_rows(T24_list) %>%
  filter_all(any_vars(!is.na(.))) %>%
  dplyr::select("ID_T24", "Fehlerhaft_T24")
```

# 1.9 Einzelteil T25

The file is imported and “Fehlerhaft” is renamed so it is clear that this column belongs to “Einzelteil T25”. Then the relevant columns were selected.

```
T25 <- read_csv("Data/Einzelteil/Einzelteil_T25.csv") %>%
  rename("Fehlerhaft_T25" = Fehlerhaft) %>%
  dplyr::select("ID_T25", "Fehlerhaft_T25")
```

# 1.10 Zulassungen

The file is imported and two column names are renamed to be identical with the primary keys. Then the relevant columns were selected.

```
Zulassungen <- read_csv2("Data/Zulassungen/Zulassungen_alle_Fahrzeuge.csv") %>%
  rename("ID_Fahrzeug" = IDNummer) %>%
  rename("Gemeinde" = Gemeinden) %>%
  dplyr::select(c(2:4))
```

# 1.11 Geodaten Gemeinden

The file is imported and the relevant columns were selected. “Postleitzahl” have a four and five digit format. For uniformity, the four-digit postal codes are adjusted by adding a 0 to the front. In addition, the name of one “Gemeinde” is missing. With the coordinates and reverse_geocode, the missing municipality name was searched and inserted into the missing entry.

```
Geodaten <- read_csv2("Data/Geodaten/Geodaten_Gemeinden_v1.2_2017-08-22_TrR.csv") %>%
  dplyr::select(c(3:6))

Geodaten$Postleitzahl <- sprintf("%05d", (Geodaten$Postleitzahl))

colSums(is.na(Geodaten))
```

```
## Postleitzahl     Gemeinde  Laengengrad  Breitengrad
##            0            1            0            0
```

```
Geodaten_NA <- which(is.na(Geodaten$Gemeinde))
Koordinaten <- Geodaten[Geodaten_NA, c("Breitengrad", "Laengengrad")]

Gemeindename <-  Koordinaten %>%
  reverse_geocode(Breitengrad, Laengengrad, method = 'osm', full_results = TRUE) %>%
  dplyr::select(municipality) %>%
  separate(municipality, "Gemeinde")

Geodaten$Gemeinde[Geodaten_NA] <- str_to_upper(Gemeindename$Gemeinde)
```

# 2. Final data set

The dataset was already tidied after the import in task 1 so that only the relevant columns are imported. In the following all tables are joined with left_join. Then only the affected vehicles were selected. A vehicle is always considered to have failed if an installed individual part, an installed component or the entire vehicle is marked as defective. In addition, the final dataset is saved in a csv file and all the files that are no longer needed are deleted.

```
Finale_Daten <- Fahrzeuge %>%
  left_join(Teile_Fahrzeuge, by = "ID_Fahrzeug") %>%
  drop_na(ID_Schaltung) %>%
  left_join(Komp_K3AG1, by = "ID_Schaltung") %>%
  left_join(Teile_Komp_K3AG1, by = "ID_Schaltung") %>%
  left_join(T21, by = "ID_T21") %>%
  left_join(T24, by = "ID_T24") %>%
  left_join(T25, by = "ID_T25") %>%
  left_join(Zulassungen, by = "ID_Fahrzeug") %>%
  left_join(Geodaten, by = "Gemeinde") %>%
  filter(Fehlerhaft_Fahrzeug == 1 | Fehlerhaft_Komp == 1 | Fehlerhaft_T21 == 1 | Fehler
haft_T24 == 1 | Fehlerhaft_T25 == 1) %>%
  mutate(Fahrzeug_betroffen = ifelse(Fehlerhaft_Fahrzeug == 1 | Fehlerhaft_Komp == 1 |
Fehlerhaft_T21 == 1 | Fehlerhaft_T24 == 1 | Fehlerhaft_T25 == 1, 1, 0)) %>%
  relocate(Fahrzeug_betroffen, .after = ID_Fahrzeug)

write.csv(Finale_Daten, "Final_dataset_group_24.csv")

rm(list = setdiff(ls(), "Finale_Daten"))
```

To display an interactive graph and heat map in the Shiny app, the sum of affected vehicles per municipality and by time must be calculated. The information is stored in a new dataset.

```
Anzahl_betroffen_Jahr <- Finale_Daten %>%
  group_by(Gemeinde, format(Zulassung, "%Y"), Postleitzahl, Laengengrad, Breitengrad) %
>%
  summarise(Anzahl_betroffen_Jahr = n()) %>%
  rename(Zulassungsjahr = `format(Zulassung, "%Y")`) %>%
  ungroup()

write.csv(Anzahl_betroffen_Jahr, "Anzahl_betroffen_Jahr.csv")
```

# 3. Evaluation

The aim of the project was to determine the defective vehicles, whereby a vehicle is defective if the vehicle itself is defective, but also if components or individual parts were defective. In this project, the vehicle types 11 and 12 with the gear “K3AG1” of the car brand “OEM1” was considered.

## 3.1 Registration of defective vehicles in municipalities over time

The interactive graph shows that the data for the affected vehicles was recorded from 2009 to 2016. The number of defective vehicles varies by municipality, e.g. in the municipality of Brey there are only a few defective vehicles, while in Koeln there are many. The main damage areas can be viewed in the heat map. If several municipalities over the entire period are selected, then you can see that the number of defective vehicles has remained approximately constant. Accordingly, problem did not occur more or less frequently over time. Except for two cases, namely in the municipalities of Hagen and Oberhausen2. In the case of Hagen, there are only two data points, which is why they are clearly visible in the change in the heat map. In Oberhausen2, on the other hand, there was a sharp increase in the number of defective vehicles in 2015.

## 3.2 Heatmap

The heat map animates the focus of damage over the period from 2009 to 2016. Looking at the entire period, there are visual changes in the heat maps every year, but they are relatively small. The defective vehicles are distributed throughout Germany but the major damage focal points exist every year and are located in the municipalities of Koeln, Dortmund, Bochum, Dresden and Leipzig. This observation also goes along with the plot in the first tab, where it was shown that the lines are approximately constant over the entire period except for two cases.

## 3.3 Final data set

The final data set shows the underlying data set of defective vehicles. The final data set has 205187 entries, accordingly 205187 vehicles are defected. If “11” is selected as vehicle type, then you can see that the table has 169955 entries. Accordingly, 169955 defective type 11 vehicles were recorded. In contrast, there were only 35232 defective vehicles of type 12.

# 4. Result

There are three tabs in the Shiny app, where you will find three tabs that visualize the data. Below are screenshots of the Shiny app showing the defective vehicles over time, a heatmap and a table with the data.

# 4.1 Registration of defective vehicles in municipalities over time

When starting the app, the first tab “Plot” is opened directly. The graph shows on the x-axis the registration years and on the y-axis is the number of affected vehicles per year. In this tab, the user can create a graph that shows the registration history of the defective vehicles in the municipalities over time. The user can select the municipalities and the time period to the right of the graph. The selected municipalities are shown in different colors in the graph and to the left of the graphic is the legend.

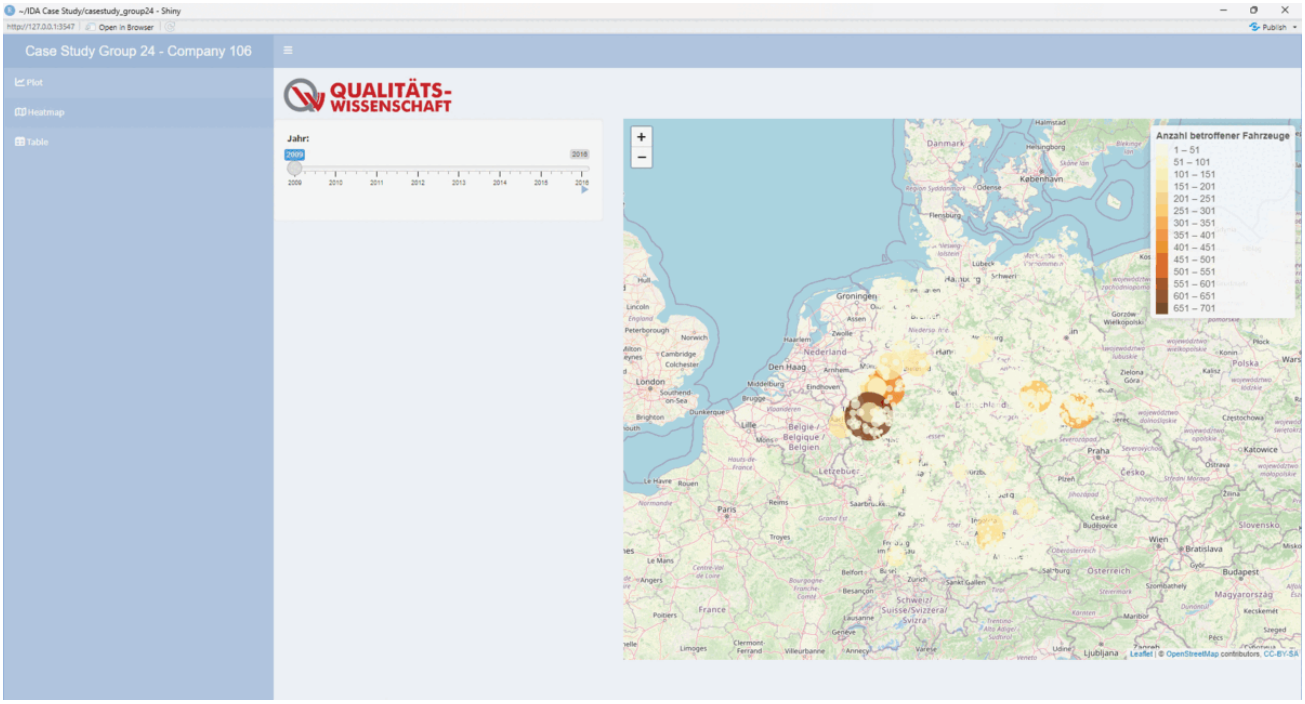
In the example below, the municipalities “Brey”, “Dortmund”, “Koeln” and “Leipzig” are selected. Their number of defective vehicles is shown from 2009 to 2016. Here you can see that there are barely any vehicles affected in the municipality of Brey. Whereas the number of affected vehicles per year in Koeln is between 630 and 700.


In the example below, the municipalities “Hagen” and “Oberhausen2” are selected. Their number of defective vehicles is shown from 2009 to 2016.

# 4.2 Heatmap

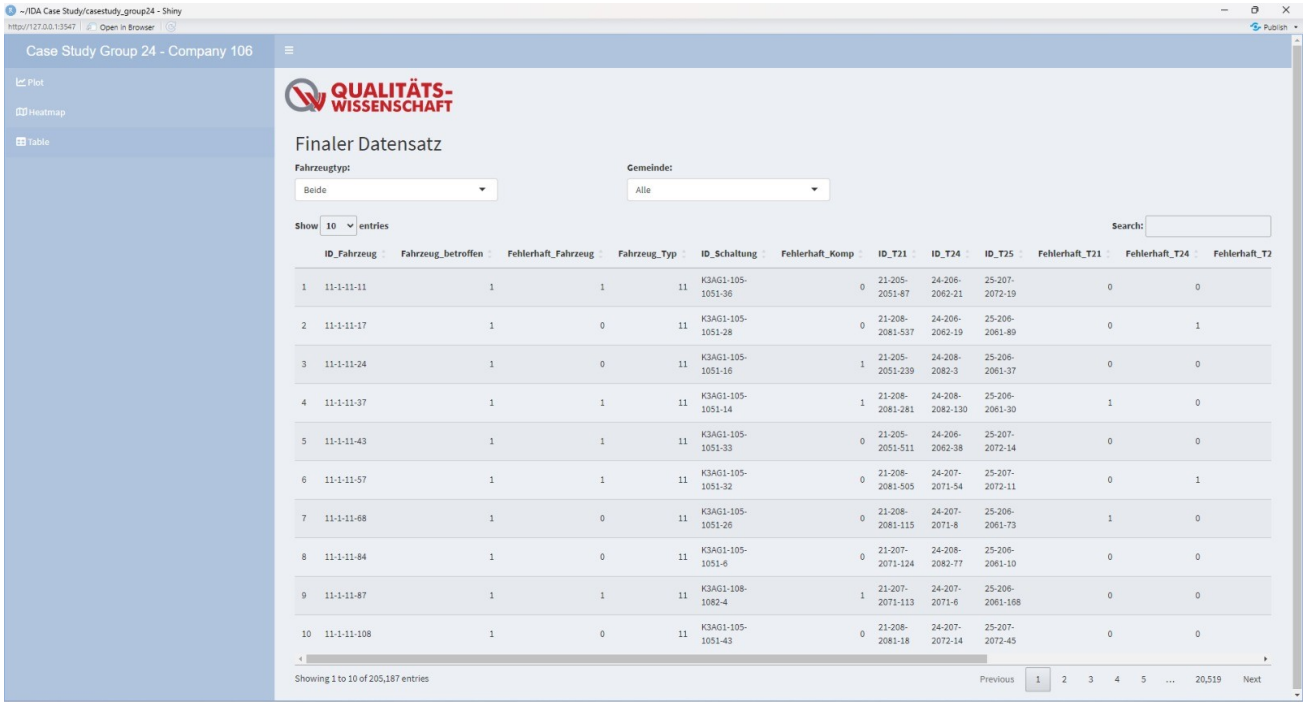
In the second tab “Heatmap”, you can find a heatmap for the visualization of focus of damages. To the left of the heatmap, the user can select the desired year. The larger the number of affected vehicles per year, the larger and darker the dot is shown in the heatmap. If the user hovers over a municipality, then further information as “Gemeinde”, “Postleitzahl” and the exact number of defective cars in that municipality is shown in a popup. In order to display the temporal course of admissions in the heatmap, a button was created that covers the specified period from 2009 to 2016. In Addition, in the upper right corner of the heatmap, there is a legend that specifies the colors in more detail. If the number is between 1 and 50, the first color is taken. With 51 to 100 the second color is assumed and so it goes on

In the example below, you can see several large damage points in the Ruhr area over time as they are mapped as larger and darker points.




# 4.3 Final data set

The last tab “Table” shows the underlying dataset “Finale_Daten” as a table. This table contains all the relevant information used for the graph and the heatmap. The table contains all important attributes of the defective vehicles, which are all the ID numbers and if it is defective, as well information about geodata and registrations. In addition, above the table the vehicle type, location and the number of displaying entries can be selected. Since the table is large, you need to scroll horizontally and vertically to view the whole table.


~/IDA Case Study/casestudy_group24 - Shiny
http://127.0.0.1:3547
Open in Browser
Publish
Case Study Group 24 - Company 106
Plot
Heatmap
Table
QUALITÄTS-WISSENSCHAFT
Finaler Datensatz
Fahrzeugtyp:
Beide
Gemeinde:
Alle
Show 10 entries
Search:
ID_Fahrzeug Fahrzeug_betroffen Fehlerhaft_Fahrzeug Fahrzeug_Typ ID_Schaltung Fehlerhaft_Komp ID_T21 ID_T24 ID_T25 Fehlerhaft_T21 Fehlerhaft_T24 Fehlerhaft_T2
1 11-1-11-11 1 1 11 K3AG1-105-1051-36 0 21-205-2051-87 24-206-2062-21 25-207-2072-19 0 0
2 11-1-11-17 1 0 11 K3AG1-105-1051-28 0 21-208-2081-537 24-206-2062-19 25-206-2061-89 0 1
3 11-1-11-24 1 0 11 K3AG1-105-1051-16 1 21-205-2051-239 24-208-2082-3 25-206-2061-37 0 0
4 11-1-11-37 1 1 11 K3AG1-105-1051-14 1 21-208-2081-281 24-208-2082-130 25-206-2061-30 1 0
5 11-1-11-43 1 1 11 K3AG1-105-1051-33 0 21-205-2051-511 24-206-2062-38 25-207-2072-14 0 0
6 11-1-11-57 1 1 11 K3AG1-105-1051-32 0 21-208-2081-505 24-207-2071-54 25-207-2072-11 0 1
7 11-1-11-68 1 0 11 K3AG1-105-1051-26 0 21-208-2081-115 24-207-2071-8 25-206-2061-73 1 0
8 11-1-11-84 1 0 11 K3AG1-105-1051-6 0 21-207-2071-124 24-208-2082-77 25-206-2061-10 0 0
9 11-1-11-87 1 1 11 K3AG1-108-1082-4 1 21-207-2071-113 24-207-2071-6 25-206-2061-168 0 0
10 11-1-11-108 1 0 11 K3AG1-105-1051-43 0 21-208-2081-18 24-207-2072-14 25-207-2072-45 0 0
Showing 1 to 10 of 205,187 entries
Previous 1 2 3 4 5 ... 20,519 Next