Part III: Visualizations in R

TMA4268 Statistical Learning V2019. Module 1: INTRODUCTION TO STATISTICAL LEARNING

Martina Hall, Michail Spitieris, Stefanie Muff, Department of Mathematical Sciences, NTNU

January 06, 2020

Contents

Introduction **1**
Packages needed 1
Data sets 2

Scatter Plot **2**

Histogram **3**

Box-plot **4**

All pairs and different plots **5**

Area chart **6**

Heat map **7**

Acknowledgements **8**

Introduction

For each of the plots (scatter plot, histogram, boxplot, area chart, heat map, correlogram) *explain what you see (including what is on the x- and y-axis) and try to transform what you see into insight about the data.* All except the correlogram use `ggplot2` for plotting. If you want to read more about the idea behind `ggplot2` (grammar of graphics) Chapter 3 of R for Data Science is a good read. Other resources are:

http://t-redactyl.io/blog/2016/03/creating-plots-in-r-using-ggplot2-part-9-function-plots.html, https://ggplot2.tidyverse.org/reference/

Packages needed

```
install.packages("car")
install.packages("faraway")
install.packages("ggplot2")
install.packages("GGally")
install.packages("reshape")
install.packages("corrplot")
install.packages("corrgram")
```

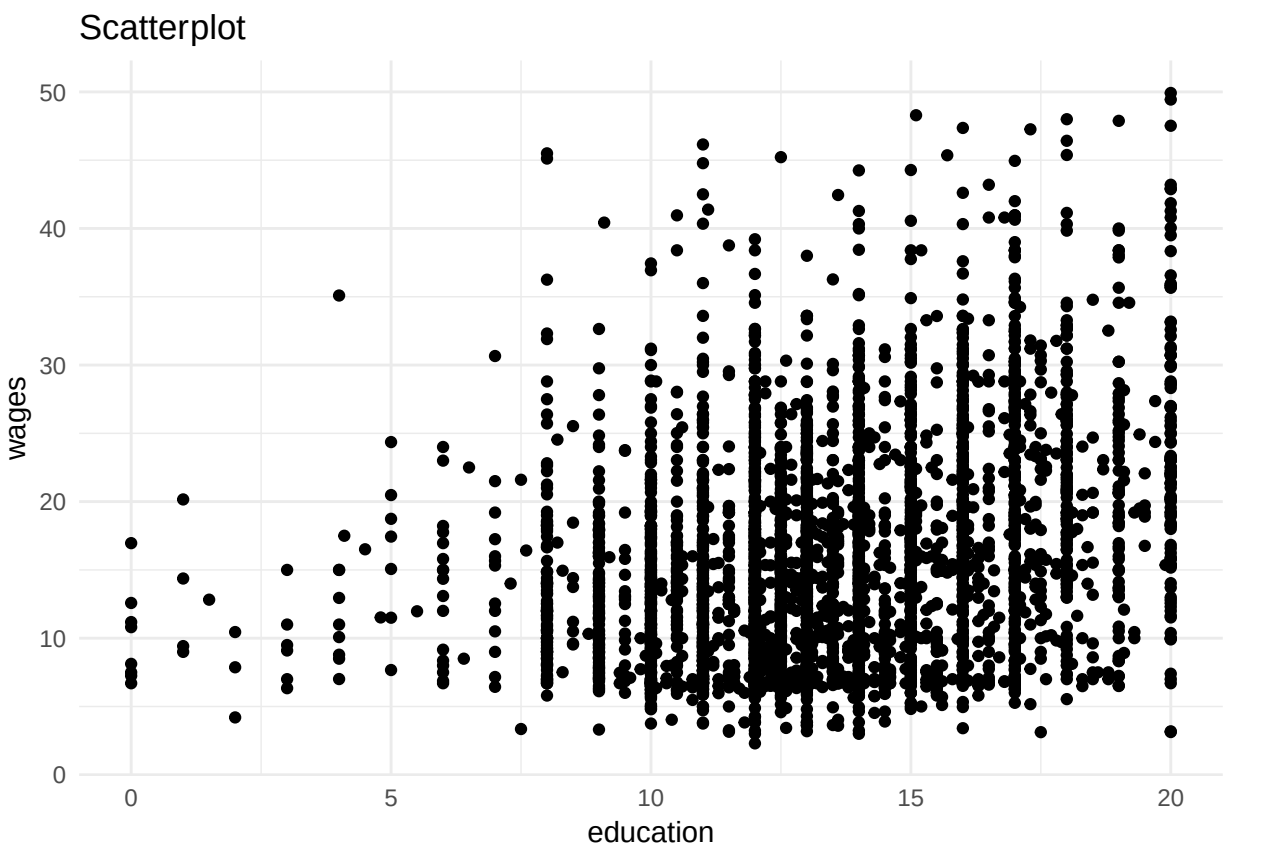
Data sets

Three different data sets are used - read descriptions in R:

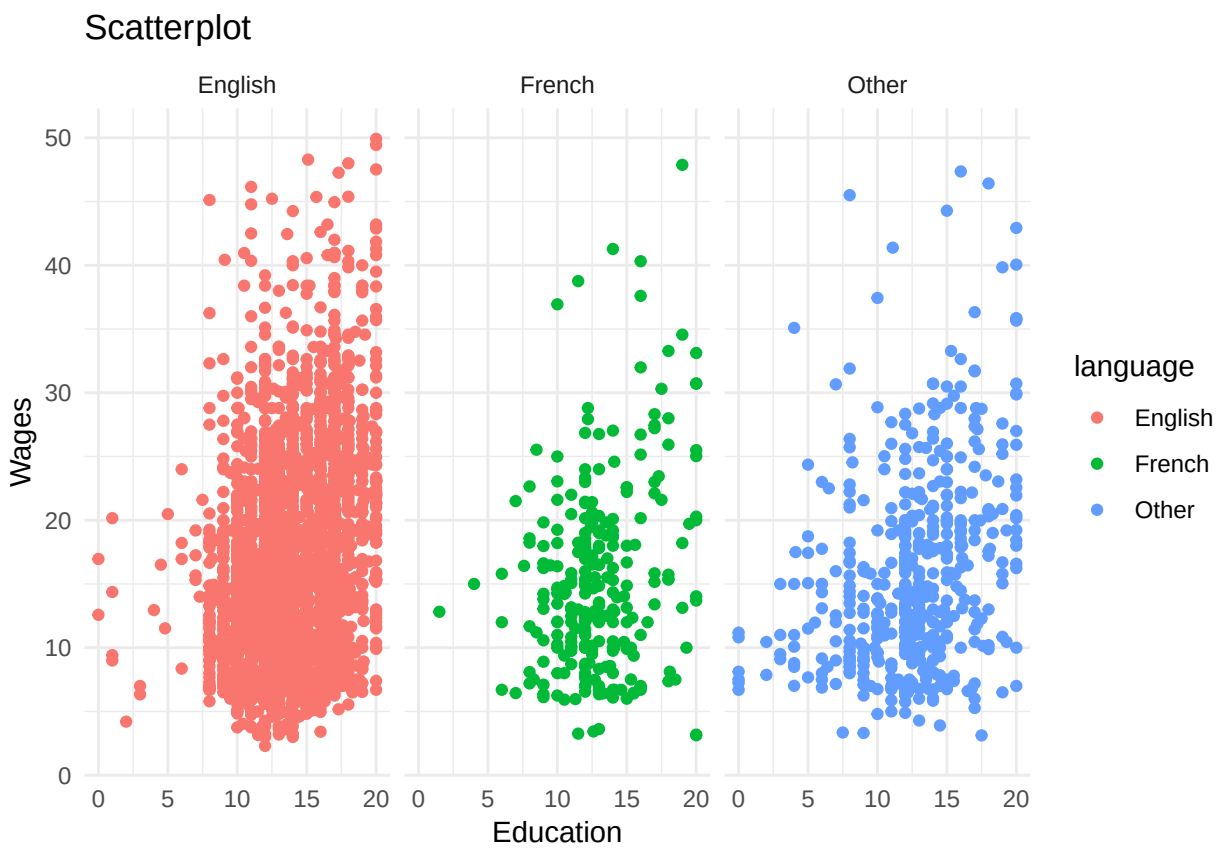
- SLID: `?car::SLID`
- mtcars: `?datasets::mtcars`
- ozone: `?faraway::ozone`

Scatter Plot

```
library(car)
library(ggplot2)
SLID = na.omit(SLID)
ggplot(SLID, aes(education, wages))+geom_point()+labs(title="Scatterplot")+theme_minimal()
```

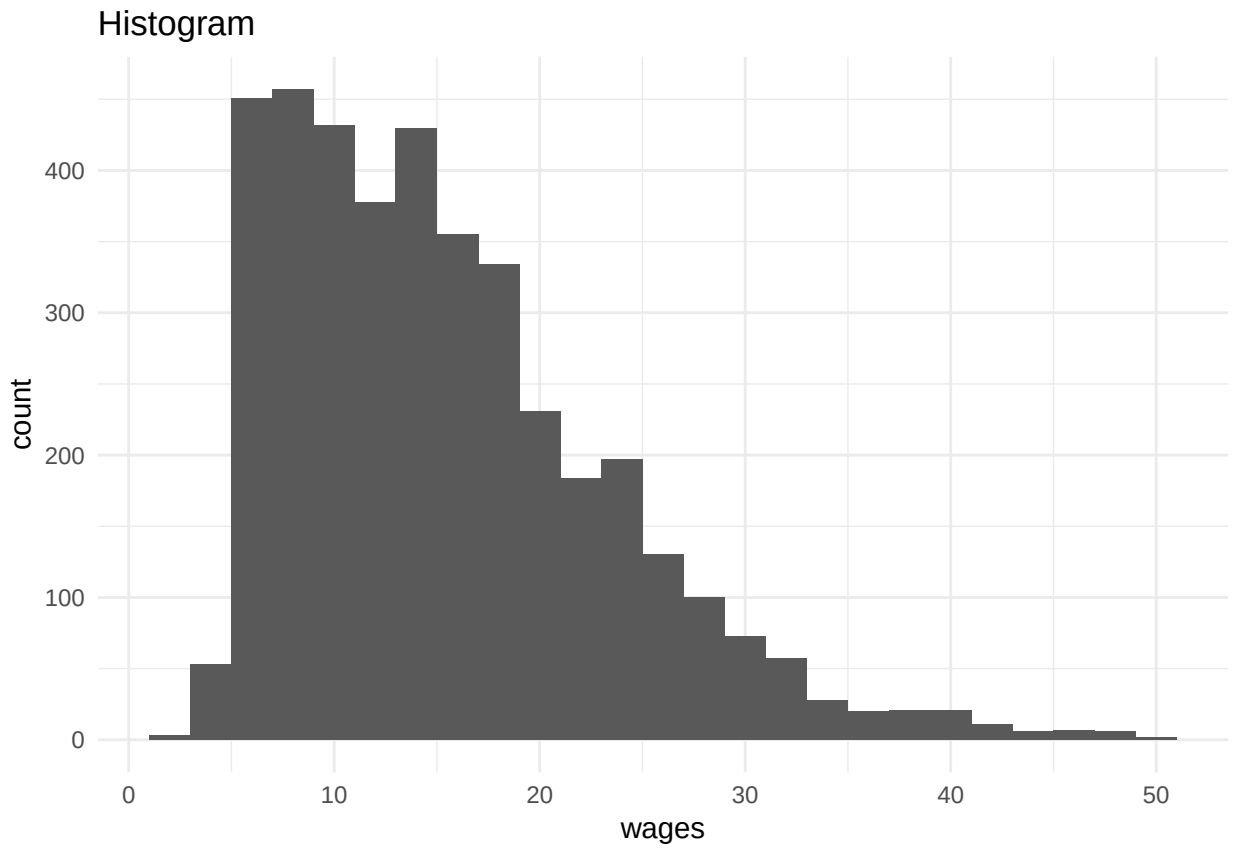



```
ggplot(SLID, aes(education, wages)) + geom_point(aes(color = language)) +
  scale_x_continuous("Education")+
  scale_y_continuous("Wages")+
  theme_bw() + labs(title="Scatterplot") + facet_wrap( ~ language)+theme_minimal()
```



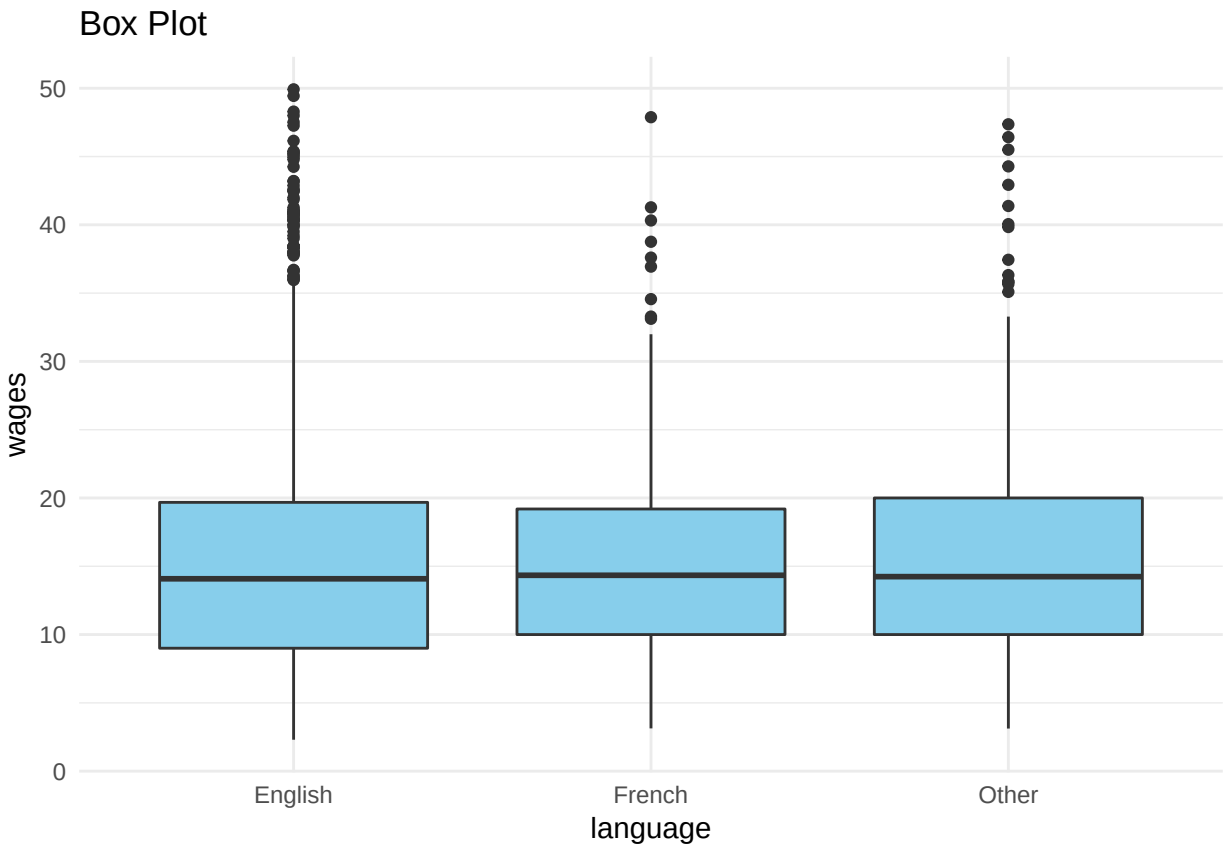
Histogram

```
ggplot(SLID, aes(wages))+geom_histogram(binwidth=2)+labs(title="Histogram")+theme_minimal()
```



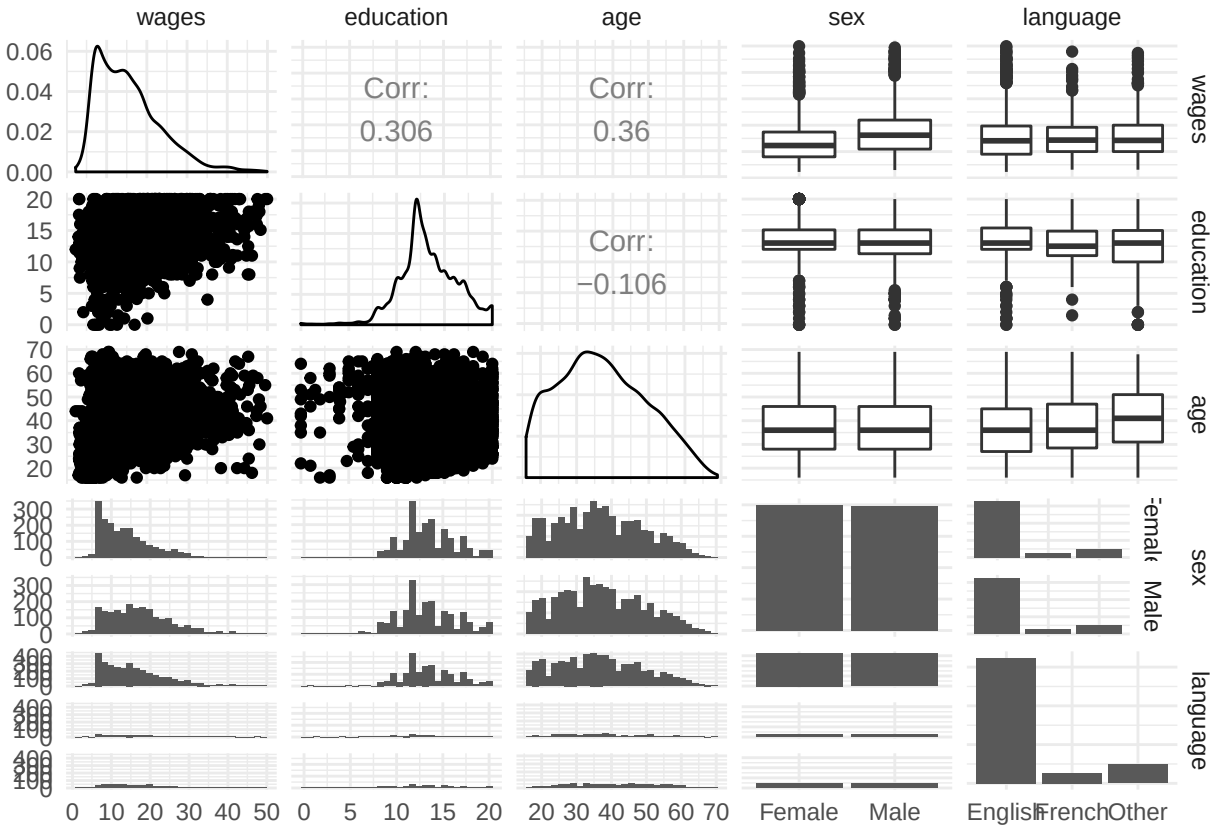
Box-plot

```
ggplot(SLID, aes(language,wages ))+geom_boxplot(fill="skyblue")+labs(title="Box Plot")+theme_minimal()
```



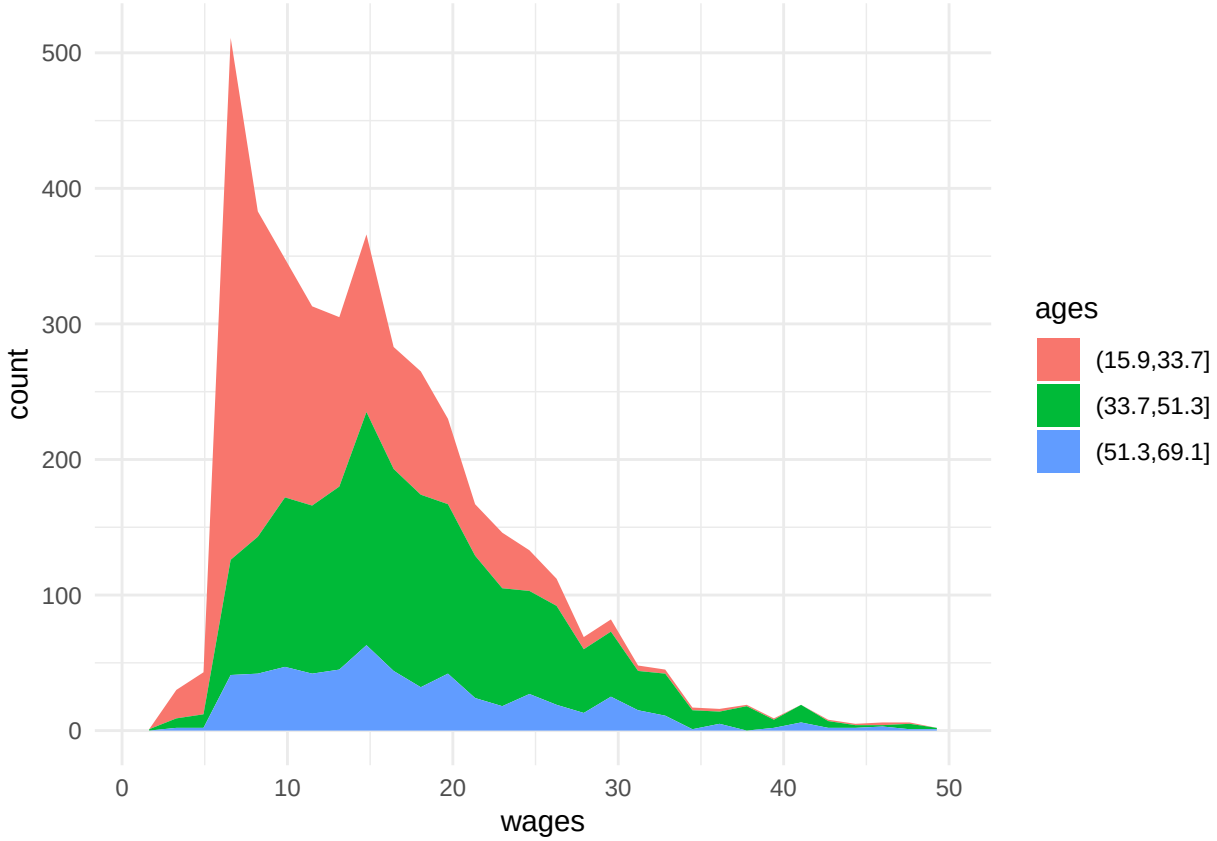
All pairs and different plots

```
library(GGally)
ggpairs(SLID)+theme_minimal()
```



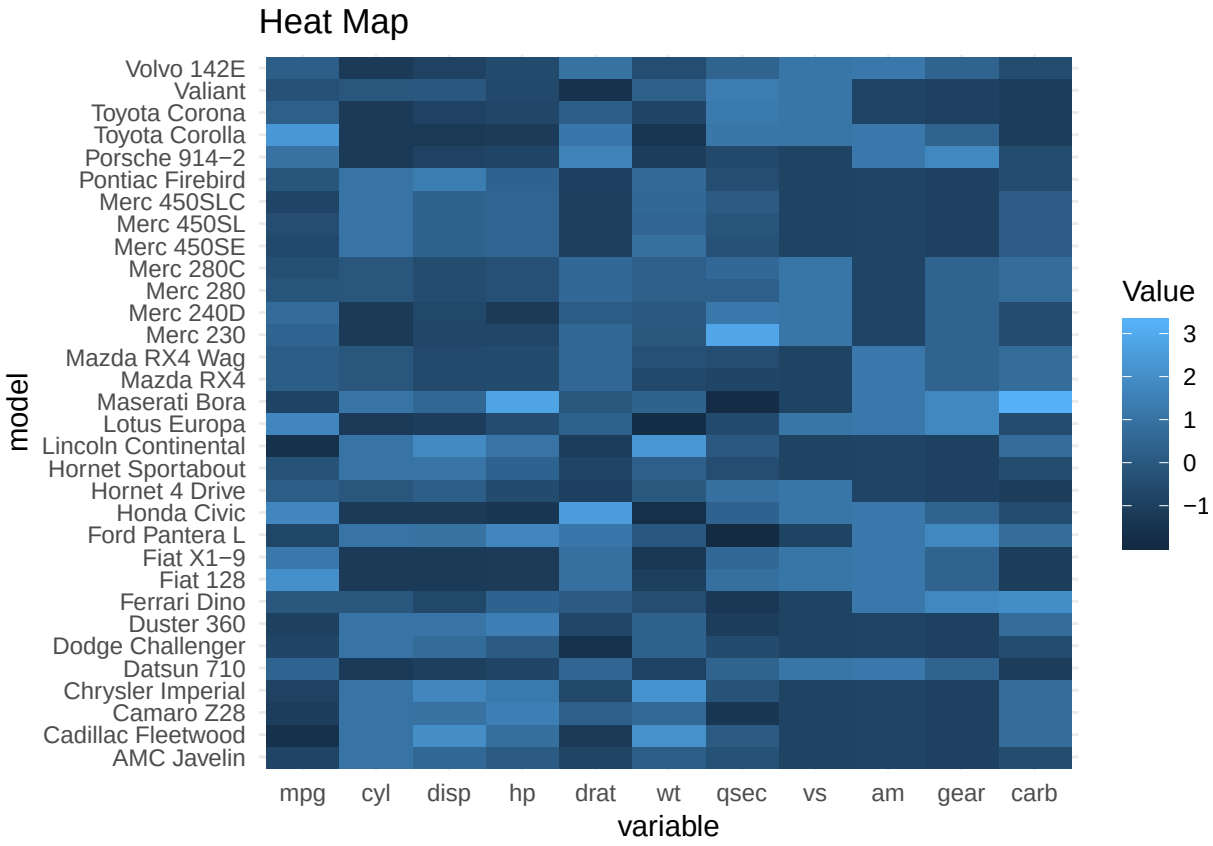
Area chart

```
ages = cut(SLID$age, breaks=3)
SLID2 = cbind(SLID, ages)
ggplot(SLID, aes(x=wages, fill=ages))+geom_area(stat="bin")+theme_minimal()
```



Heat map

```
library(reshape)
head(mtcars)
carsdf = data.frame(scale(mtcars))
carsdf$model = rownames(mtcars)
cars_melt = melt(carsdf, id.vars="model")
ggplot(cars_melt, aes(x =variable, y = model)) +
  geom_raster(aes(fill=value)) +
  labs(title="Heat Map") +
  scale_fill_continuous(name="Value") +
  theme_minimal()
```




```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```


Acknowledgements

We thank Mette Langaas and her PhD students from 2018 and 2019 for building up the original version of this sheet.