Data Exploration in R

Heatmaps

Prof. Dr. Jan Kirenz

HdM Stuttgart

Heatmap

- In this presentation you will learn how to perform data exploration using heatmaps.

You can use this kind of heatmap if the variable is:

- Categorical,
- discrete with only a few values or
- continuous

Data

- We use the dataset wage
- It contains wage and other data for a group of 3000 male workers.
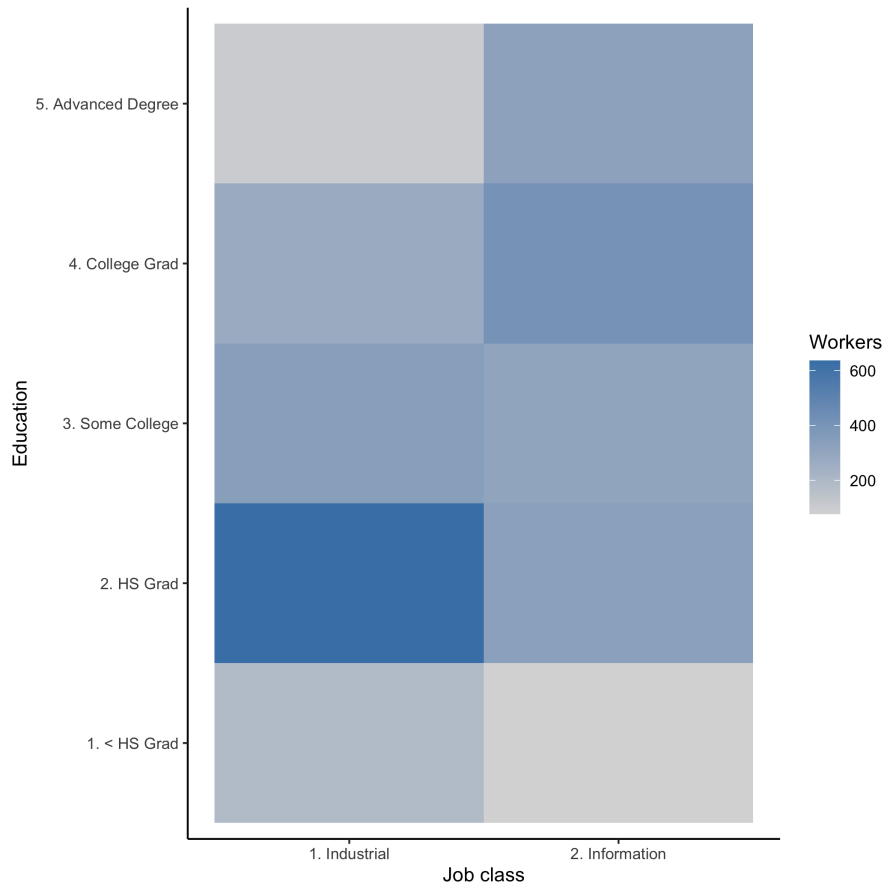
```
library(tidyverse)

wage_df <- read_csv("https://raw.githubusercontent.com/kirenz/datasets/master/wage.csv")
```

- The data includes 3000 observations on 11 variables.

Two categorical variables

- Heatmaps for categorical and discrete Variables
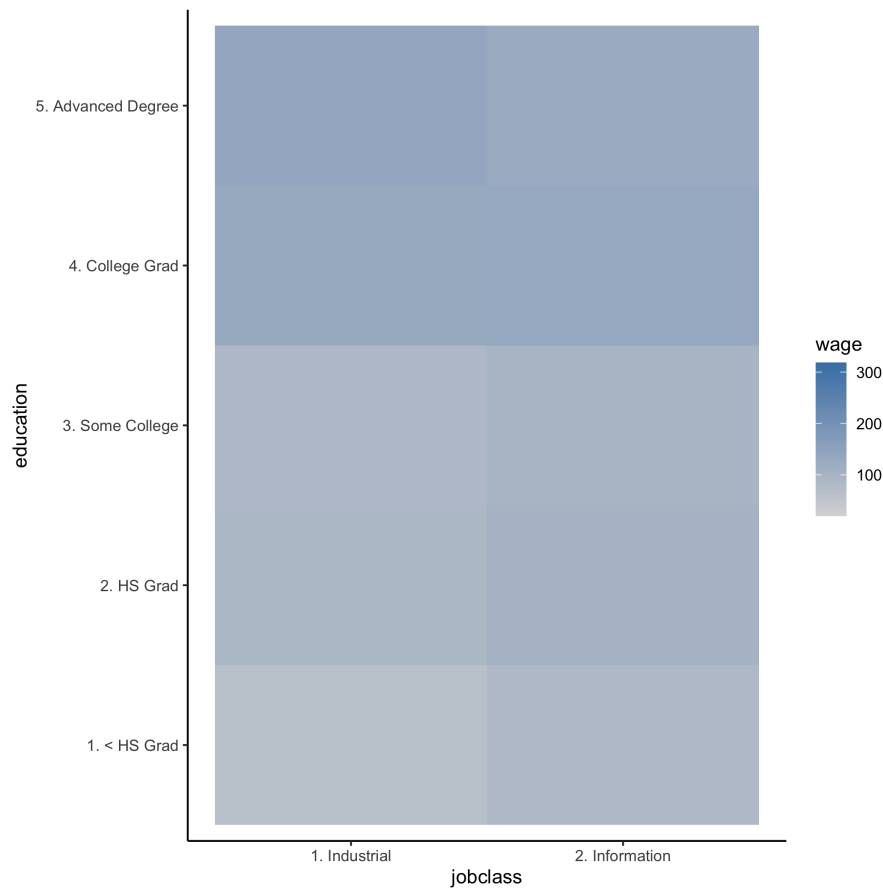- We use the function geom_bin2d() from the ggplot2 package.

Two categorical variables

```
wage_df %>%
  ggplot(aes(jobclass, education)) +
  geom_bin2d() +
  scale_fill_gradient(low = "gray85", high = "steelblue") +
  labs(fill = "Workers", y = "Education", x = "Job class") +
  theme_classic()
```

- geom_bin2d() divides the plane into rectangles, counts the number of cases in each rectangle, and then (by default) maps the number of cases to the rectangle's fill.

Combination of Variables

- Combination of categorical and continous variables

Data Exploration in R | Prof. Dr. Jan Kirenz

6 / 8

Combination of Variables

- Combination of categorical and continous variables

```
wage_df %>%
  ggplot(aes(jobclass, education, fill = wage )) +
  geom_tile() +
  scale_fill_gradient(low = "gray85", high = "steelblue")  +
  theme_classic()
```

Thanks!

Prof. Dr. Jan Kirenz

HdM Stuttgart
Nobelstraße 10
70569 Stuttgart