CaseStudy2018

Ilse Beelen, Floor Komen and Lotte Pater

14 december 2018

```r
setwd("~/Studie/Statistics & Data Science/Semester 1/Linea
#GetData

Data <- read.csv("1_clean_data/voting_and_demographics.csv"
                 stringsAsFactors=F, header = T)



#Remove and rename columns
Data <- Data[,-15]
colnames(Data) <- c("Muni", "VVD","CDA","PVV", "D66", "SP",

# Non_west_perc is not linear. Therefore, we create a dummy
# Level 1: x < 5%
# Level 2: 5 <= x < 10%
# Level 3: x >= 10%
Data$Non_west <- ifelse(Data$Non_west_perc < 0.05, 1, NA)
Data$Non_west <- ifelse(Data$Non_west_perc >= 0.05
                        & Data$Non_west_perc < 0.1, 2, Data
Data$Non_west <- ifelse(Data$Non_west_perc >= 0.1, 3, Data$
```

Introduction and data summary

Introductie later

- Parties to research: CDA and GroenLinks
- Demographics: Amount of 60+ residents, Non-west residents, mean income, highly educated residents

Data visualisation CDA

Correlation between explanatory & respons variables

Data visualisation GroenLinks

Data Cleaning

Formulate model

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Final model

```
plot(pressure)
```