

Übung 14

Explorative Datenanalyse und Visualisierung

Wintersemester 2019
S. Döhler (FBMN, h_da)

Name: Kessler & Cisternas Seeger

Punkte:

Aufgabe 30. *Arbeiten Sie mit dem Datensatz `USArrests` im R-Basispaket.*

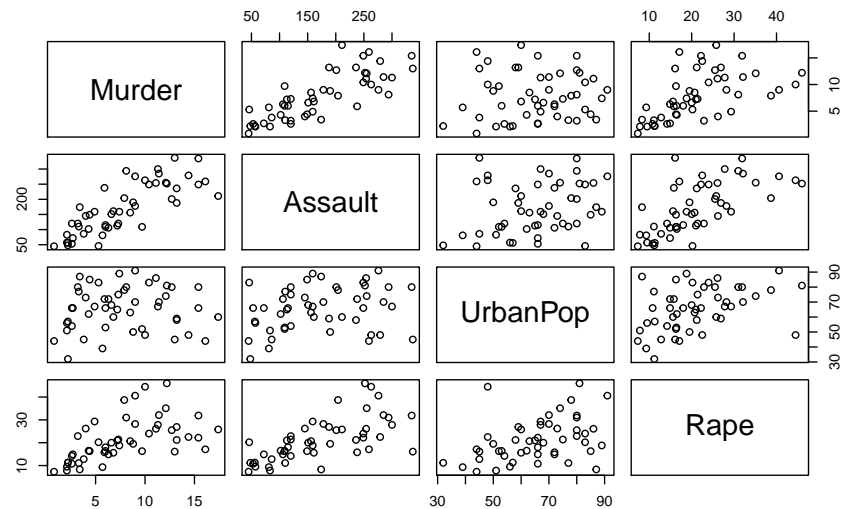
- Machen Sie sich mit dem Datensatz vertraut.*
- Erzeugen Sie mit `ggplot2` einen scatterplot mit den Daten (x -Achse: Urban-Pop, y -Achse: verschiedene Verbrechenhäufigkeiten) sowie der jeweiligen Regressionen*
- Erzeugen Sie mit `ggplot2` Histogramme der verschiedene Verbrechenhäufigkeiten.*
- Sortieren Sie die Daten, so dass Sie pro Verbrechensart nach Häufigkeiten sortiert sind.*

Lösung Aufgabe a

```
df = USArrests
summary(df)
```

#>	Murder	Assault	UrbanPop	Rape
#> Min.	: 0.800	Min. : 45.0	Min. : 32.00	Min. : 7.30
#> 1st Qu.	: 4.075	1st Qu.: 109.0	1st Qu.: 54.50	1st Qu.: 15.07
#> Median	: 7.250	Median : 159.0	Median : 66.00	Median : 20.10
#> Mean	: 7.788	Mean : 170.8	Mean : 65.54	Mean : 21.23
#> 3rd Qu.	: 11.250	3rd Qu.: 249.0	3rd Qu.: 77.75	3rd Qu.: 26.18
#> Max.	: 17.400	Max. : 337.0	Max. : 91.00	Max. : 46.00

```
plot(df)
```

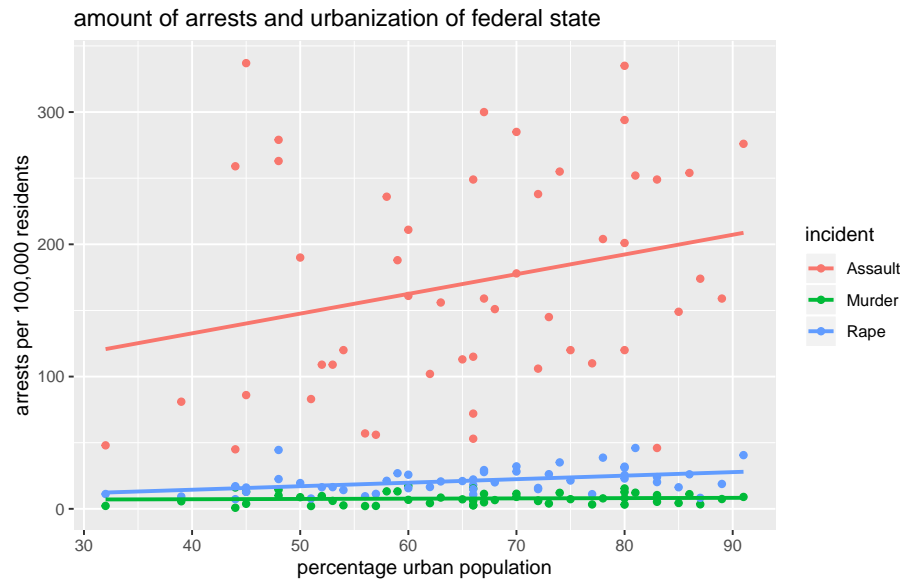


Tidy the data:

```
library(tidyverse)
pdf <- pivot_longer(data=df, cols = c("Murder","Assault","Rape"), names_to = "incident", val
```

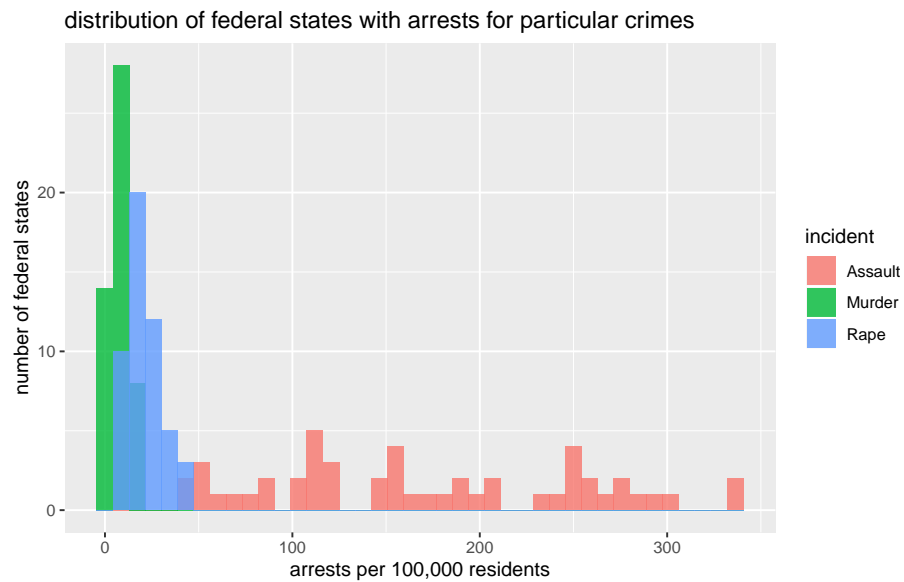
Aufgabe b

```
library(ggplot2)
ggplot(data=pdf, aes(x = UrbanPop, y = amount, color = incident)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(x = "percentage urban population") +
  labs(y = "arrests per 100,000 residents") +
  labs(title = "amount of arrests and urbanization of federal state")
```



Aufgabe c

```
ggplot(pdf, aes(x = amount, fill= incident)) +  
  geom_histogram(alpha=0.8, position="identity", bins = 40) +  
  labs(y = "number of federal states") +  
  labs(x = "arrests per 100,000 residents") +  
  labs(title = "distribution of federal states with arrests for particular crimes")
```



Aufgabe d

```
pdf %>% arrange(incident, amount, .by_group = TRUE)
#> # A tibble: 150 x 3
#>   UrbanPop incident amount
#>   <int> <chr>    <dbl>
#> 1     44 Assault     45
#> 2     83 Assault     46
#> 3     32 Assault     48
#> 4     66 Assault     53
#> 5     57 Assault     56
#> 6     56 Assault     57
#> 7     66 Assault     72
#> 8     39 Assault     81
#> 9     51 Assault     83
#> 10    45 Assault     86
#> # ... with 140 more rows
```

Anmerkungen/Korrektur _____