

# Data Mining

Vincent Dahmen 6689845      Rafael Heid 6704828

15. April 2016

## 1

1.  $A=[1;2;3]$  oder  $A=[1:3]'$
2.  $B(2:3,:)$
3.  $\text{ones}(3,1) * 5$
4.  $E=[B \ D]$
5.  $\text{zeros}(2,3)$  oder  $[0 \ 0; \ 0 \ 0; \ 0 \ 0]$ 
  - (a)  $M(1,:)$
  - (b)  $M(:,5)$
  - (c)  $M([2:2:\text{end}],:)$
  - (d)  $M(:,[1:2:\text{end}])$

## 2

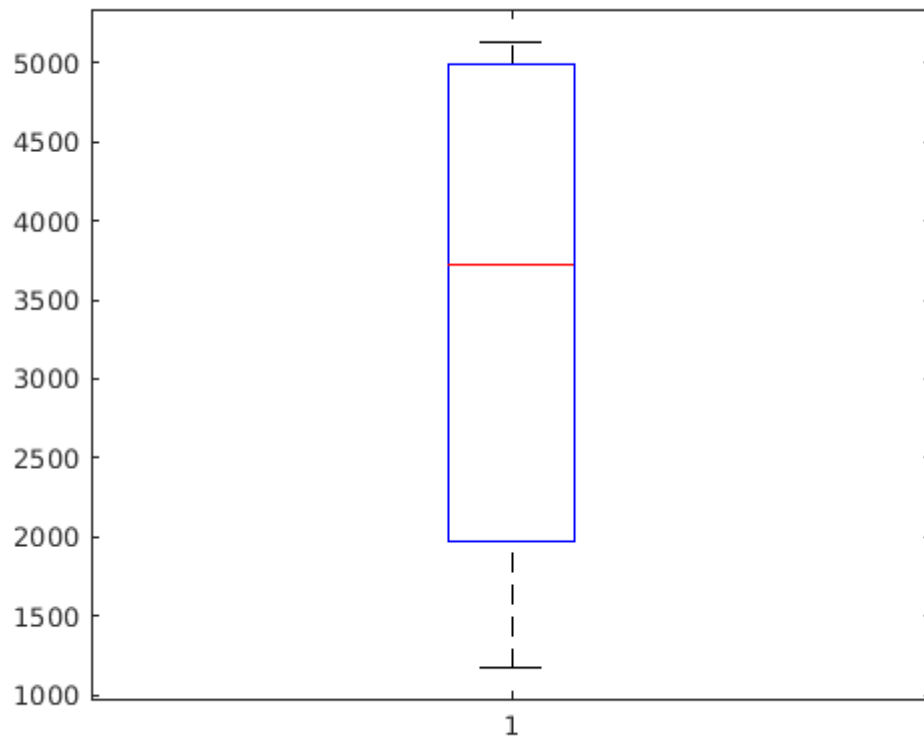
Nominale Daten

- Hunderassen
- Automarken
- Städtenamen

Ordinale Daten

- Monate
- gewichtsbeschreibende Adjektive
- Sonnenstand

### 3



1. Die Figur (..) zeigt das Boxplot-Diagramm von income1

- Daten für ein Diagramm müssen in einer Spalte liegen
- Daten müssen nicht sortiert sein
- Min- und Max-Werte sind jeweils durch den oberen bzw unteren Querstrich dargestellt
- Der Durschnitt (mean) ist der mittlere Abstand der äußeren Striche
- Der Median (also das 50.Perzentil) wird durch den roten Querstrich innerhalb der Box dargestellt
- Q1 und Q3 sind jeweils Anfang bzw Ende der blauen Box
- um ein Boxplot-Diagramm zu erstellen folgt man folgenden Schritten:

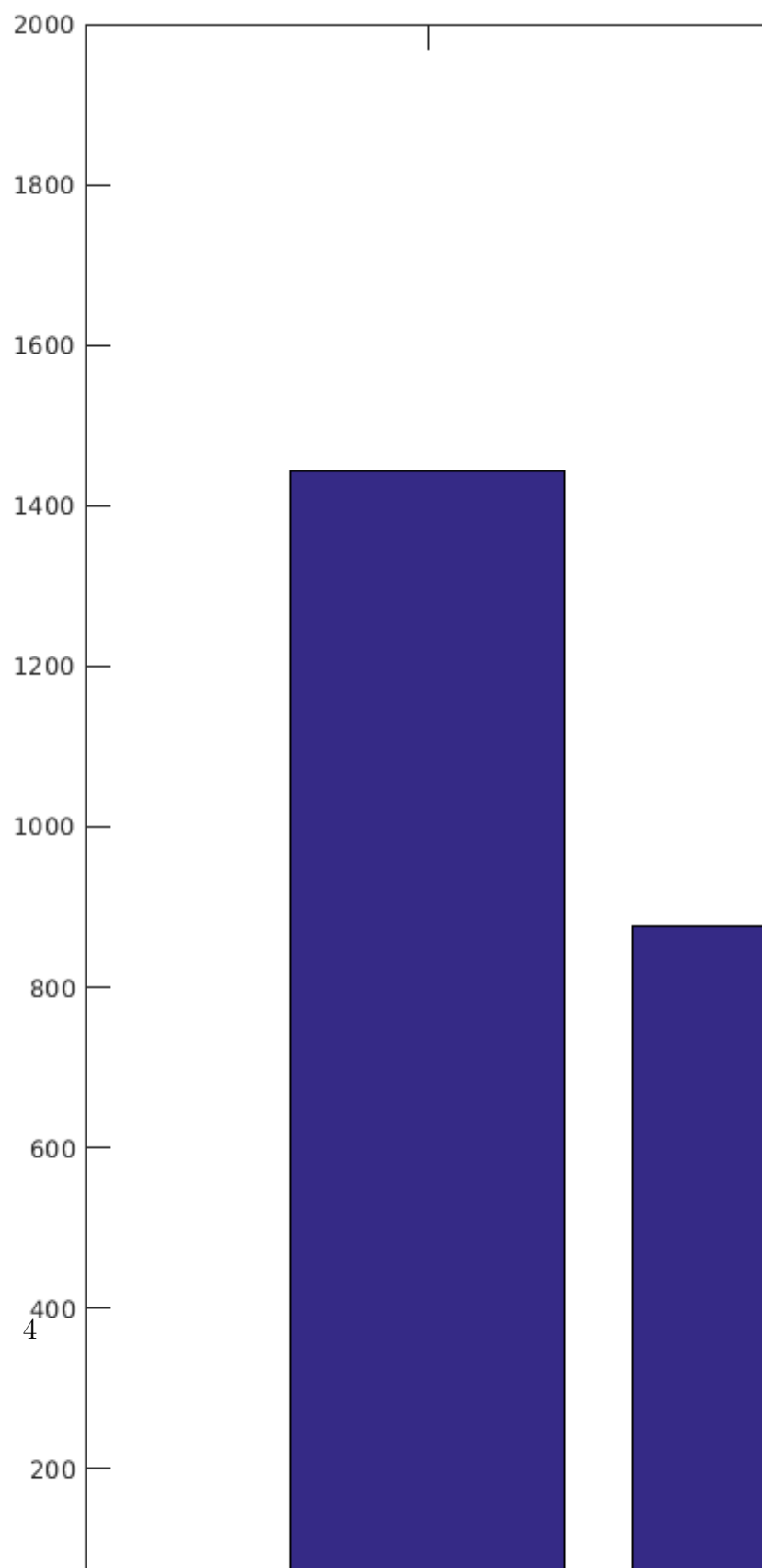
(a) figure(1)

(b) `boxplot(income1)`

2. Die Figur (..) zeigt den Vergleich von `income1` und `income2`
  - Da sich die Anzahl der Datensätze verändert hat ist die Box (leicht) verschoben (also Q1 und Q3)
  - aus den gleichen Gründen hat sich auch der Median verschoben
  - das rote Kreuz ist ein Ausreißer der sehr weit von dem Rest abweicht
3. Um vor Ausreißern gefeit zu sein empfiehlt sich die Betrachtung des Medians bei dichten Wertemengen

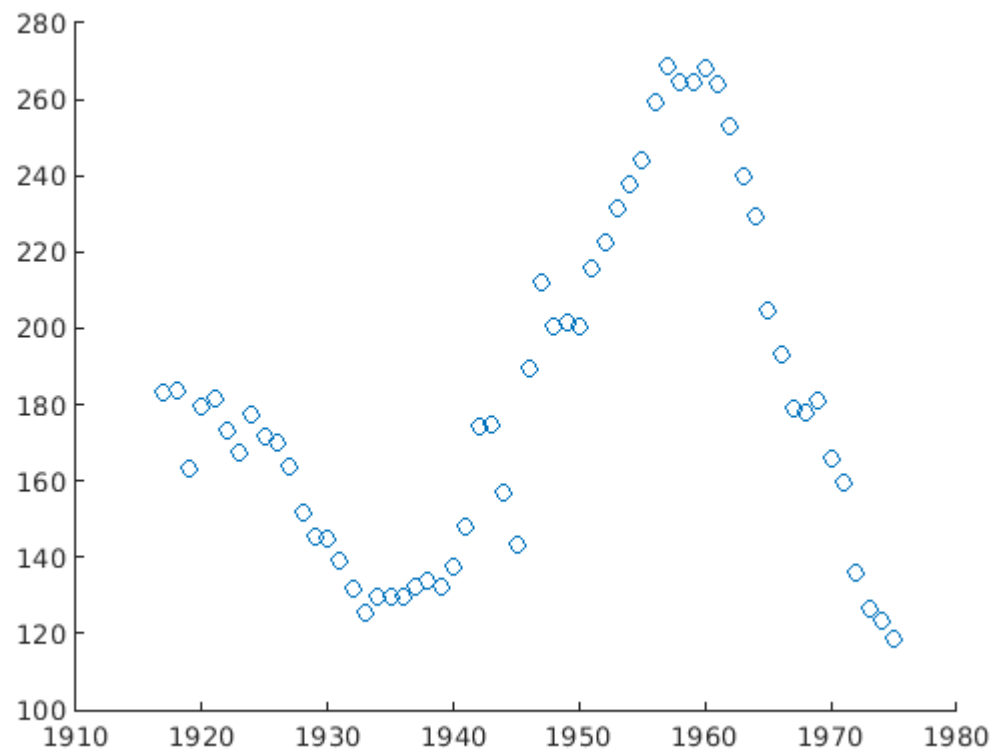
## 4

1. Die folgenden Punkte beziehen sich auf das Kreisdiagramm mit den Wahlergebnissen
  - Beschriftung ist irreführend, da die Summe der angegebenen Verteilung  $\neq$  100 %
2. Die folgenden Punkte beziehen sich auf das Balkendiagramm mit den Pokalen
  - Die als 'Balken' gewählten Pokale sind schlecht zu vergleichen
  - Es wird lediglich die Anzahl der gewonnen, nicht aber der gespielten Spiele angegeben
  - Mangelnde Details zum Verlauf der Spiele
3. Die folgenden Punkte beziehen sich auf die verbleibenden Diagramme
  - Die vermeintlich zusammenpassenden Kurven liegen in verschiedenen Zeiten



2. die Mindestlöhne illustrieren wir mithilfe eines Bar-Charts. Dafür sprechen folgende Punkte:

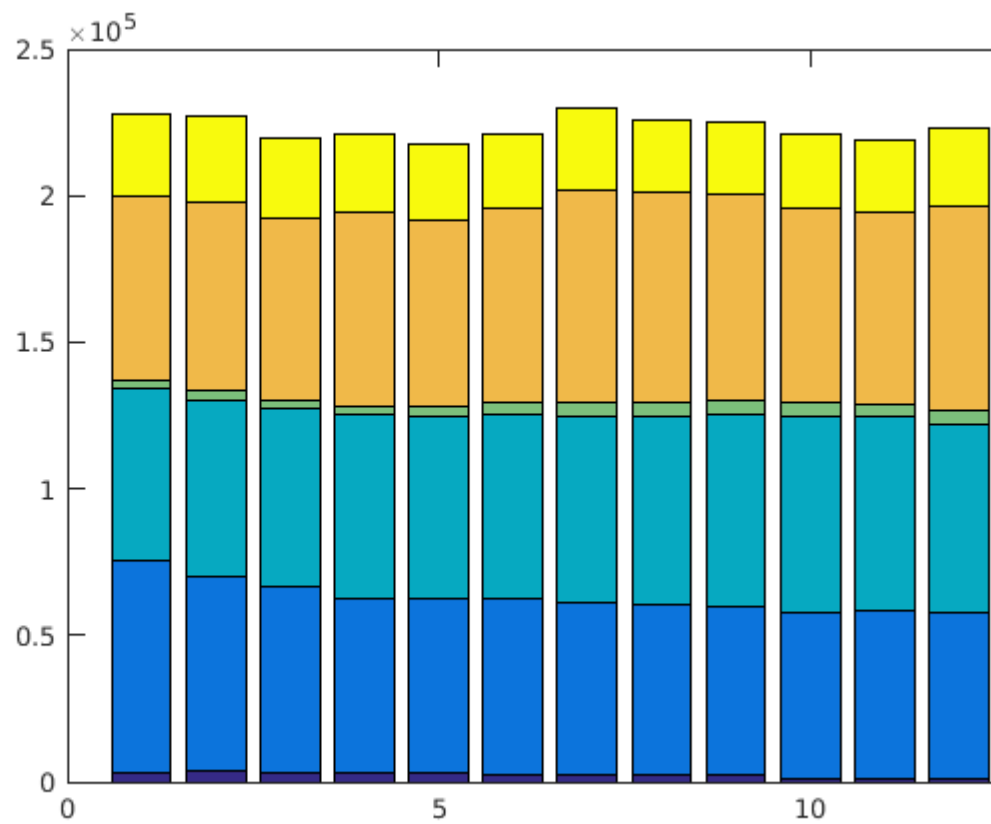
- es werden nur wenige verglichen
- die Werte sind ähnlich
- es besteht ein Interesse an dem direkten Vergleich der Werte



3.

4. die Geburtsrate stellen wir als Scatter-Plot da. Dafür sprechen folgende Punkte:

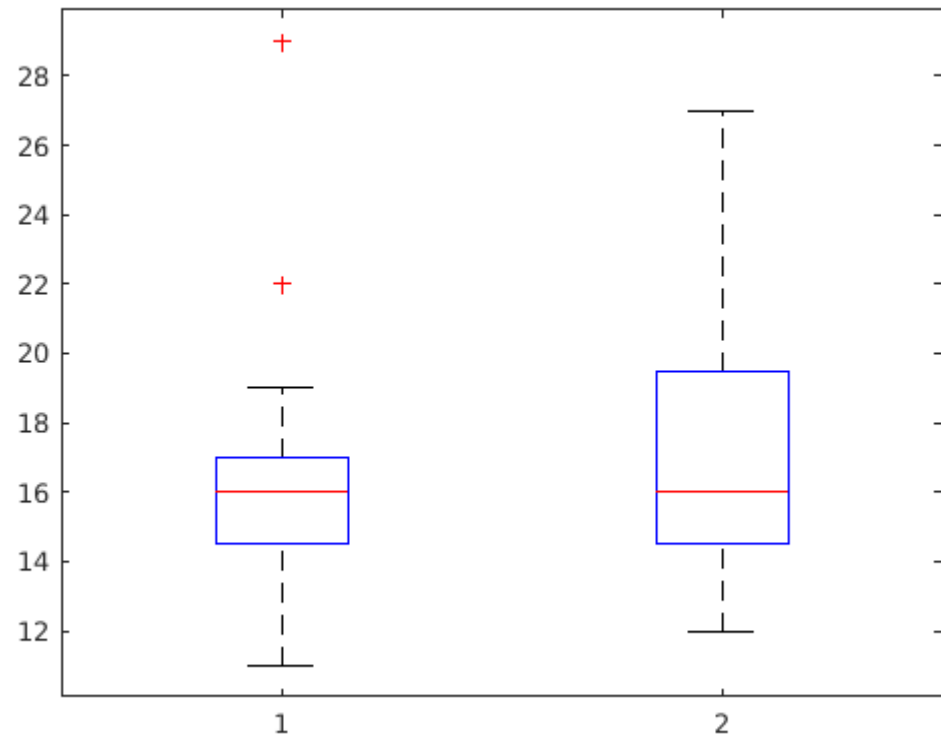
- die Daten sind als Tupel mit Zeitpunkt verfügbar
- es besteht ein Interesse an der Entwicklung
- die Datensätze sind standatisiert



5.

6. Die Energieverbrauchsstatistik geben wir als Bar-Chart an

- die 6 Spalten einer Zeile werden immer für eine Balken benutzt und farbig getrennt
- jeder Balken (jede Datenzeile) steht für ein Jahr
- dadurch lässt sich die Verteilung mehrerer Jahre gut vergleichen

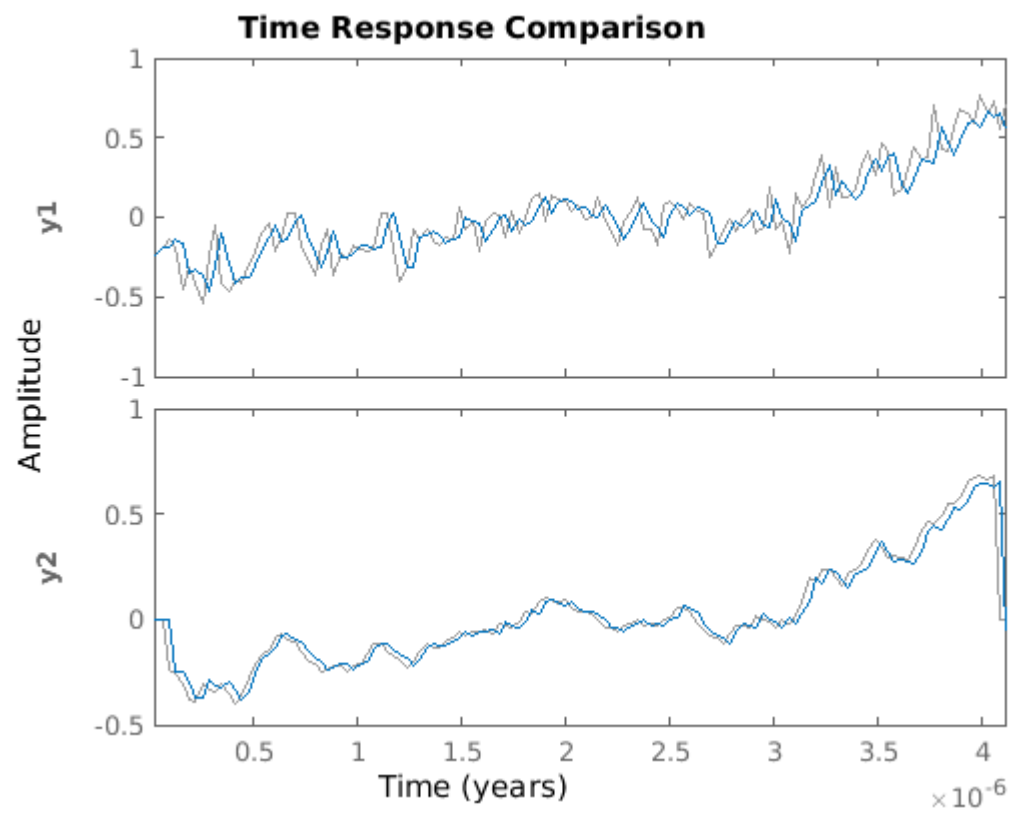


7.

8. Für die Visualisierung des Stroop nehmen Boxplots

- zwei Datenreihen, die miteinander verglichen werden
- gleiches Experiment
- mehrere Ergebnisse

9. Das Klima wird als timeseries dargestellt



10.