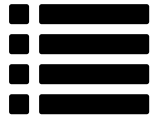
Reproduzierbare Analysen

Schweizer Statistiktage 2017

Andrea Schnell
Statistisches Amt Kanton Zürich

Agenda



 Reproduzierbarkeit

 RMarkdown

 Tools

 RCommunity

Reproduzierbarkeit
Konzept

> «... you can't write down mouse clicks»

Reproduzierbarkeit
Konzept

> «... you can't write down mouse clicks»

> the data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding.

Reproduzierbarkeit
Konzept

> «... you can't write down mouse clicks»

> the data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding.

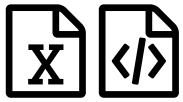
- nachvollziehbar
- wiederverwendbar
- automatisierbar

Reproduzierbarkeit

Analyseschritte

- Daten

- Auswertungen



- Grafiken

- Text

- Publikation

Reproduzierbarkeit
Analyseschritte

- Daten
- Auswertungen
- Grafiken
- Text
- Publikation

- viele Dokumente
- händische, manuelle Schritte
- schwierig nachvollziehbar
- kaum reproduzierbar

RMarkdown

Idee

RMarkdown

Idee

- Gesamte Analyse inkl. Resultaten an einem Ort
- Kommentare, R Code, Metadaten, Grafiken
- Outputformate: Word, PDF, HTML
- Layoutvorgaben in Templates (LaTeX, CSS)

RMarkdown

Idee

- Gesamte Analyse inkl. Resultaten an einem Ort
- Kommentare, R Code, Metadaten, Grafiken
- Outputformate: Word, PDF, HTML
- Layoutvorgaben in Templates (LaTeX, CSS)

> - nachvollziehbar
> - reproduzierbar
> - vereinfacht Zusammenarbeit

RMarkdown
Output PDF

```
# Bevölkerung im Kanton Zürich

## Per Jahresende
Ende `r bev %>% filter(STATYEAR == max(STATYEAR)) %>% select(STATYEAR) %>% distinct()
%>% .$STATYEAR` waren`r bev %>% filter(STATYEAR == max(STATYEAR)) %>% summarise(n =
sum(BEV_TOTAL)) %>% .$n %>% format(., big.mark = "'") ` Personen im Kanton Zürich
wohnhaft.

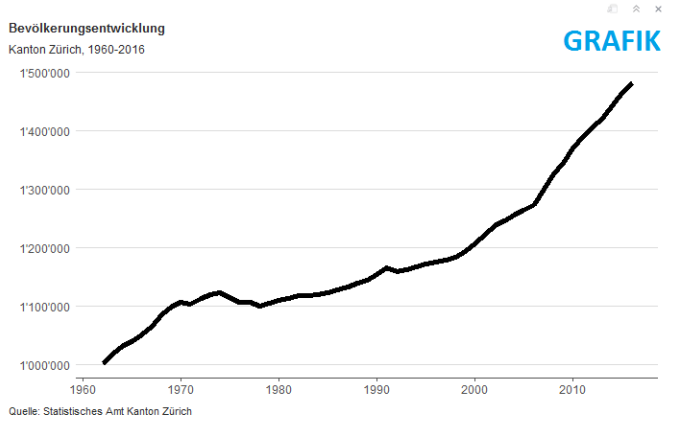
## Zeitverlauf
```

TEXT

CODE

```
```{r, echo = F}
g <- bev %>% group_by(STATYEAR) %>% summarise(n = sum(BEV_TOTAL)) %>%
  ggplot(., aes(x = STATYEAR, y = n))+
  geom_line(size = 2)+
  theme_stat_tex()+
  scale_y_continuous(labels = format_format(big.mark = "'"))+
  labs(x =NULL, y = NULL)

flush_plot(g, title = "Bevölkerungsentwicklung", subtitle = "Kanton Zürich, 1960-2016",
           caption = "Quelle: Statistisches Amt Kanton Zürich")
```
```



GRAFIK

RMarkdown
Output PDF

```
# Bevölkerung im Kanton Zürich
```

TEXT

```
## Per Jahresende
Ende `r bev %>% filter(STATYEAR == max(STATYEAR)) %>% select(STATYEAR) %>% distinct()
%>% .$STATYEAR` waren`r bev %>% filter(STATYEAR == max(STATYEAR)) %>% summarise(n =
sum(BEV_TOTAL)) %>% .$n %>% format(., big.mark = "'") ` Personen im Kanton Zürich
wohnhaft.

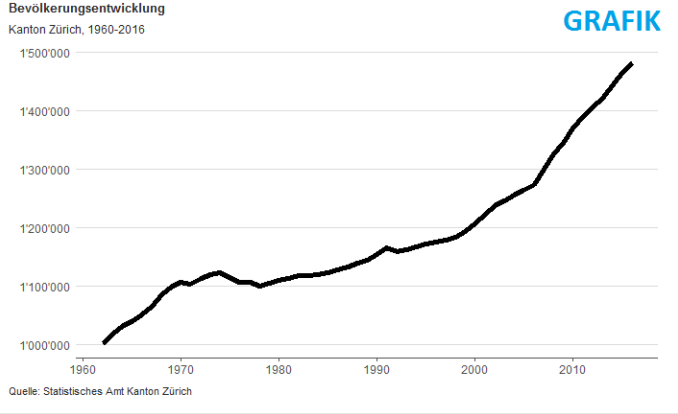
## Zeitverlauf
```

CODE

```
```{r, echo = F}
g <- bev %>% group_by(STATYEAR) %>% summarise(n = sum(BEV_TOTAL)) %>%
  ggplot(., aes(x = STATYEAR, y = n))+
  geom_line(size = 2)+
  theme_stat_tex()+
  scale_y_continuous(labels = format_format(big.mark = "'"))+
  labs(x =NULL, y = NULL)

flush_plot(g, title = "Bevölkerungsentwicklung", subtitle = "Kanton Zürich, 1960-2016",
           caption = "Quelle: Statistisches Amt Kanton Zürich")
```
```

GRAFIK



Bevölkerung im Kanton Zürich

Andrea Schnell

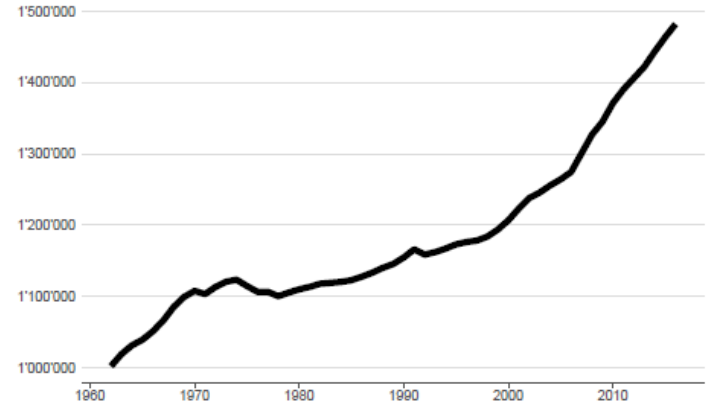
Per Jahresende

Ende 2016 waren1'482'003 Personen im Kanton Zürich wohnhaft.

Zeitverlauf

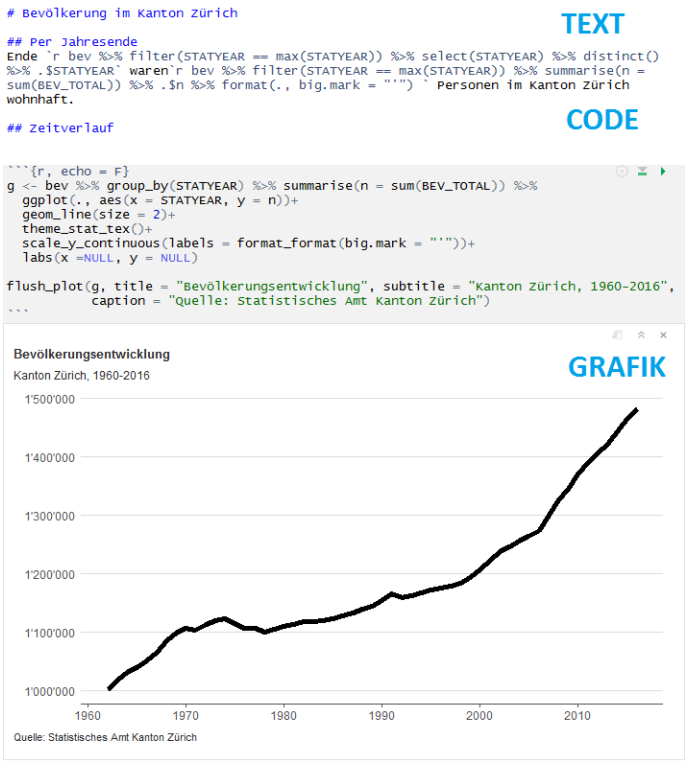
Bevölkerungsentwicklung

Kanton Zürich, 1960-2016



Quelle: Statistisches Amt Kanton Zürich

RMarkdown
Output HTML




```
# Bevölkerung im Kanton Zürich

## Per Jahresende
Ende `r bev %>% filter(STATYEAR == max(STATYEAR)) %>% select(STATYEAR) %>% distinct()
%>% .$STATYEAR` waren`r bev %>% filter(STATYEAR == max(STATYEAR)) %>% summarise(n =
sum(BEV_TOTAL)) %>% .$n %>% format(., big.mark = "'") ` Personen im Kanton Zürich
wohnhaft.

## Zeitverlauf

```{r, echo = F}
g <- bev %>% group_by(STATYEAR) %>% summarise(n = sum(BEV_TOTAL)) %>%
  ggplot(., aes(x = STATYEAR, y = n))+
  geom_line(size = 2)+
  theme_stat_tex()+
  scale_y_continuous(labels = format_format(big.mark = "'"))+
  labs(x =NULL, y = NULL)

flush_plot(g, title = "Bevölkerungsentwicklung", subtitle = "Kanton Zürich, 1960-2016",
           caption = "Quelle: Statistisches Amt Kanton Zürich")
```
```

RMarkdown
Output HTML

TEXT

```
# Bevölkerung im Kanton Zürich

## Per Jahresende
Ende `r bev %>% filter(STATYEAR == max(STATYEAR)) %>% select(STATYEAR) %>% distinct() %>% .$STATYEAR` waren`r bev %>% filter(STATYEAR == max(STATYEAR)) %>% summarise(n = sum(BEV_TOTAL)) %>% .$n %>% format(., big.mark = "'") ` Personen im Kanton Zürich wohnhaft.

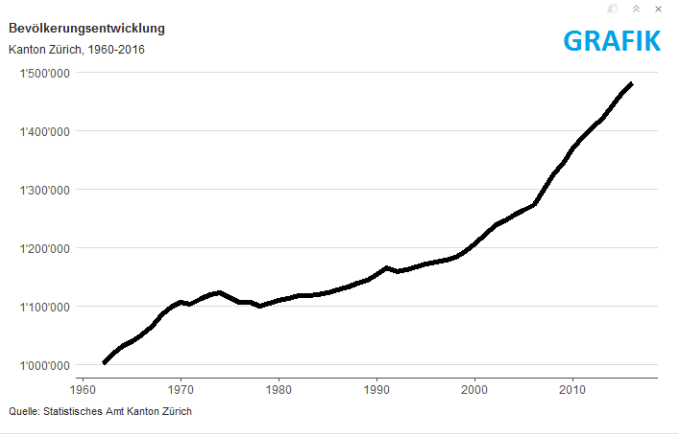
## Zeitverlauf
```

CODE

```
```{r, echo = F}
g <- bev %>% group_by(STATYEAR) %>% summarise(n = sum(BEV_TOTAL)) %>%
  ggplot(., aes(x = STATYEAR, y = n))+
  geom_line(size = 2)+
  theme_stat_tex()+
  scale_y_continuous(labels = format_format(big.mark = "'"))+
  labs(x =NULL, y = NULL)

flush_plot(g, title = "Bevölkerungsentwicklung", subtitle = "Kanton Zürich, 1960-2016",
           caption = "Quelle: Statistisches Amt Kanton Zürich")
```
```

GRAFIK



Bevölkerung

Kanton Zürich

Per Jahresende | Zeitverlauf

Ende 2016 waren1’482’003 Personen im Kanton Zürich wohnhaft.

Tools
R und mehr

- RMarkdown
- Templates
- RProjects

- RPackages

- Git/Github

Community

«... one of R's best features»

Community

Zurich R User Group

«The use of R in public administration»

Talk 1: Read and process unstructured data with R

Max Grütter, Statistical Office
Slides

Talk 2: Mastering Corporate Design with R

Andrea Schnell and Thomas Lo Russo, Statistical Office
Slides

Talk 3: Illustration of pupils' diversity with R

Flavian Imlig, Zurich State Department of Education
Slides

Talk 4: Making Public Data Public

Christoph Sax, Christoph Sax Data Analytics
Slides

https://statistikzh.github.io/RMeetup

Vielen Dank
für Ihre Aufmerksamkeit!

Andrea Schnell
andrea.schnell@statistik.ji.zh.ch

@statistik_zh

https://statistikzh.github.io/SST17

Slides created via the R package **xaringan**.