

Case-Study zur Arbeitslosigkeit in Deutschland

Organisatorische Hinweise

- + Viele Deadlines
- + Ungewohntes Format (sehr technisch)
- + Github, RStudio, R
- + Arbeitsschritte mit Github (3. Problem Set von Github herunterladen und lösen)

Dies ist alles neu und das ist uns bewusst!

Organisatorische Hinweise

- + Viele Deadlines
- + Ungewohntes Format (sehr technisch)
- + Github, RStudio, R
- + Arbeitsschritte mit Github (3. Problem Set von Github herunterladen und lösen)

Dies ist alles neu und das ist uns bewusst!

Was bieten wir ihnen:

- + Tutorium mit Lara morgen um 14:15 Uhr in BBB
- + Hilfe im Forum
- + Video Tutorials
 - + klonen von Github (mit Github-Desktop)
 - + RTutor Problem Set lokal ausführen (gleich wie auf der RStudio Cloud)

Recap letzte Vorlesungseinheit

- + Verschiedene Arten einen Datensatz einzulesen
 - + `readr`, `readxl`, `haven`...
- + Variablenbezeichnungen stehen nicht zwangsläufig in erster Spalte
- + Es gibt oft und viele NAs in echten Daten
 - + Konsistenzchecks wichtig
- + Datensätze sind nicht immer in der Form das wir diese direkt Einlesen können
 - + Aus verschiedenen Quellen einlesen, z.B. über eine `for`-Schleife
 - + Umformen, da die Daten im `wide`-Format vorliegen -> `pivot_longer`
- + Es ist wichtig sich selbst ein Bild von den Daten zu machen

Analyse der Daten

Deskriptive vs. induktive Statistik

- ✚ Deskriptive Statistik (beschreibende Statistik) ist beschreibend (wer hätte es gedacht)
- ✚ Induktive (auch schließende) Statistik versucht aus der Stichprobe auf die Grundgesamtheit zu schließen

Deskriptive vs. induktive Statistik

- ✚ Deskriptive Statistik (beschreibende Statistik) ist beschreibend (wer hätte es gedacht)
- ✚ Induktive (auch schließende) Statistik versucht aus der Stichprobe auf die Grundgesamtheit zu schließen
- ✚ Keine Unterscheidung in der Formel
- ✚ Keine Unterscheidung in dem Datensatz der verwendet wird

Deskriptive vs. induktive Statistik

- + Deskriptive Statistik (beschreibende Statistik) ist beschreibend (wer hätte es gedacht)
- + Induktive (auch schließende) Statistik versucht aus der Stichprobe auf die Grundgesamtheit zu schließen
- + Keine Unterscheidung in der Formel
- + Keine Unterscheidung in dem Datensatz der verwendet wird

| Worin genau besteht der Unterschied zwischen der deskriptiven und der induktiven Statistik?

Deskriptive Statistik

- + Beschreibung des Datensatzes
 - + Beispiel: Daten von der Agentur für Arbeit über die Arbeitslosenquote in den Landkreisen
- + Mehrere Arten denkbar
 - + Tabellenform
 - + Visualisierung mittels Schaubildern

Sie wollen etwas über ihren aktuellen Datensatz lernen

Induktive Statistik

- + Interesse gilt nicht dem Datensatz selbst, sondern der Population
 - + Sie haben keine Vollerhebung durchgeführt, sondern nur eine (zufällige) Stichprobe der Population gezogen
- + Beispiel: Mikrozensus, d.h. eine Befragung von zufällig ausgewählten Haushalten in Deutschland
- + Sie wollen aus der Stichprobe schätzen, wie sich die beobachtete Größe in der Population verhält
- + Es gibt viele Arten der induktiven Statistik. Die zwei häufigsten:
 - + Vorhersage
 - + Erkennen kausaler Zusammenhänge

Induktive Statistik

- + Interesse gilt nicht dem Datensatz selbst, sondern der Population
 - + Sie haben keine Vollerhebung durchgeführt, sondern nur eine (zufällige) Stichprobe der Population gezogen
- + Beispiel: Mikrozensus, d.h. eine Befragung von zufällig ausgewählten Haushalten in Deutschland
- + Sie wollen aus der Stichprobe schätzen, wie sich die beobachtete Größe in der Population verhält
- + Es gibt viele Arten der induktiven Statistik. Die zwei häufigsten:
 - + Vorhersage
 - + Erkennen kausaler Zusammenhänge

In die induktive Statistik tauchen wir nächstes Semester tiefer ein.

Deskriptive Statistik

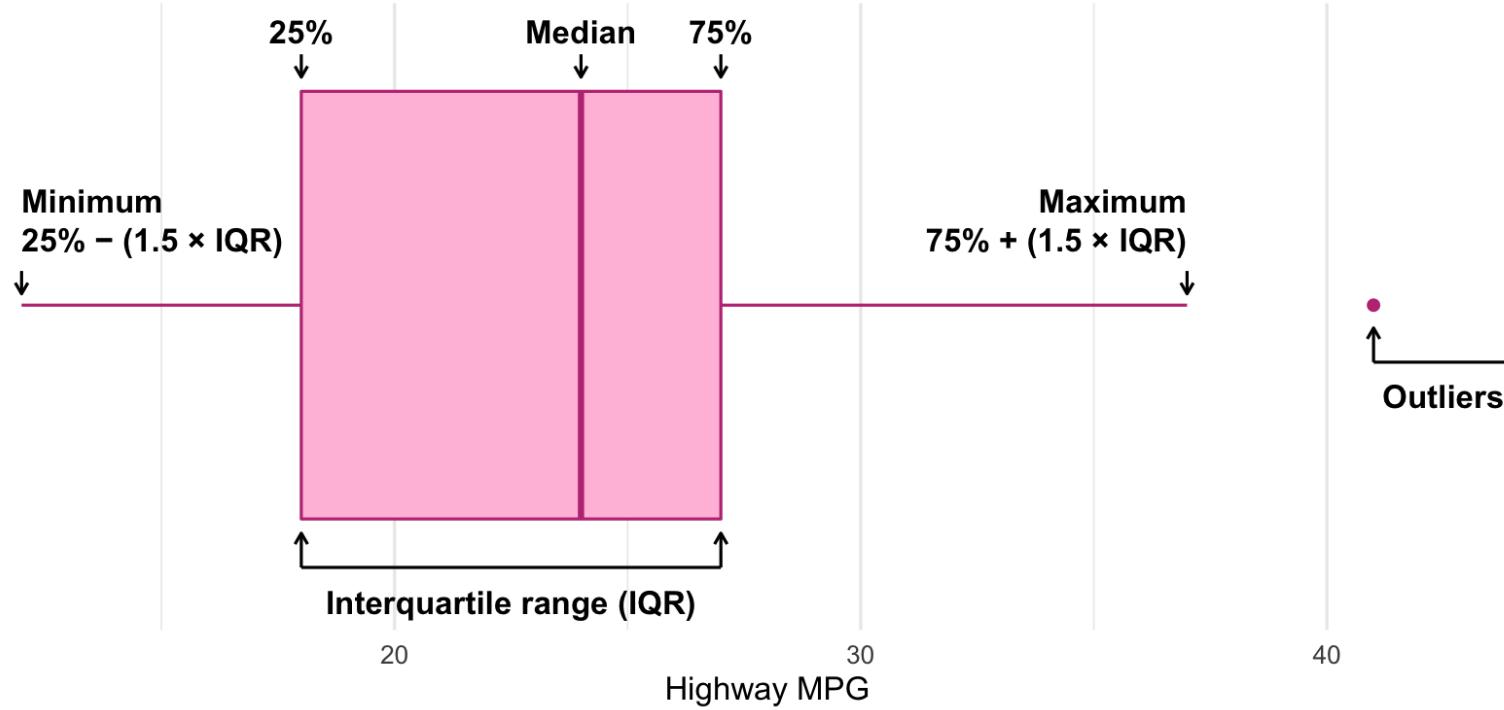
Univariate deskriptive Statistik

- + Eine Variable wird dargestellt:
 - + Verteilung
 - + Mittelwert
 - + Standardabweichung
 - + Median
 - + Quantile
- + Überblick verschaffen, Eigenschaften der Variablen aufzeigen

Univariate deskriptive Statistik

- + Darstellung über eine Tabelle
 - + Median, Mittelwert, Standardabweichung und Quantile
- + Darstellung über einen Boxplot
 - + Median, Inter-Quartile-Range (ICR), Ausreißer
- + Darstellung über ein Histogram
 - + Verteilung mit Anzahl an Beobachtungen
- + Darstellung über einen Kerndichteschätzer
 - + Verteilung mit Dichte

Univariate deskriptive Statistik (Boxplot)



Bivariate deskriptive Statistik

Darstellung von Zusammenhängen zweier Variablen

- + Korrelation zweier Variablen
- + Wenn sich eine Variable verändert, wie verändert sich die andere Variable?

Darstellung als:

- + Streudiagramm
- + Korrelationskoeffizient (meist innerhalb eines Korrelationsmatrix)

Wie sieht die deskriptive Statistik in der
Praxis aus?

Zweiter Teil der Case Study

Eingelesene Daten deskriptiv untersuchen

- + Erster Schritt: Deskriptive Tabellen mit `kableExtra` und `gt`
- + Zweiter Schritt: Grafiken mit `ggplot2`

Zweiter Teil der Case Study

Eingelesene Daten deskriptiv untersuchen

- + Erster Schritt: Deskriptive Tabellen mit `kableExtra` und `gt`
- + Zweiter Schritt: Grafiken mit `ggplot2`

Ziele des zweiten Teils der Case Study:

- + Daten visualisieren und Zusammenhänge grafisch veranschaulichen
- + Deskriptive Analysen mittels Korrelationstabellen und deskriptiven Tabellen anfertigen
- + Das Verständnis wie Sie ihre Informationen zu bestimmten Fragestellungen möglichst effektiv aufbereiten
- + Interaktive Grafiken erstellen

Zweiter Teil der Case Study

Eingelesene Daten deskriptiv untersuchen

- + Erster Schritt: Deskriptive Tabellen mit `kableExtra` und `gt`
- + Zweiter Schritt: Grafiken mit `ggplot2`

Ziele des zweiten Teils der Case Study:

- + Daten visualisieren und Zusammenhänge grafisch veranschaulichen
- + Deskriptive Analysen mittels Korrelationstabellen und deskriptiven Tabellen anfertigen
- + Das Verständnis wie Sie ihre Informationen zu bestimmten Fragestellungen möglichst effektiv aufbereiten
- + Interaktive Grafiken erstellen

Im dritten RTutor Problem Set werden Sie Visualisierung zu einzelnen Ländern auf europäischer Ebene.

Daten und Pakete laden

Wir laden die aus Teil 1 erstellten Datensätze:

```
library(tidyverse)
library(skimr)
library(sf)
library(viridis)
library(plotly)
library(kableExtra)
library(gt)
library(corr)
```

```
# Daten einlesen
einkommen <- readRDS("../case-study/data/einkommen.rds")
bundesland <- readRDS("../case-study/data/bundesland.rds")
landkreise <- readRDS("../case-study/data/landkreise.rds")
bip_zeitreihe <- readRDS("../case-study/data/bip_zeitreihe.rds")
gemeinden <- readRDS("../case-study/data/gemeinden.rds")
gesamtdaten <- readRDS("../case-study/data/gesamtdaten.rds")
schulden_bereinigt <- readRDS("../case-study/data/schulden_bereinigt.rds")
```

Deskriptive Analysen

Arbeitslosenquote berechnen

Zuerst: Überblick über die Daten gewinnen

- + Wie viele Landkreise haben wir in den Daten?
- + Wie ist die Verteilung der Schulden, Arbeitsenquote und des BIP?

Arbeitslosenquote berechnen

Zuerst: Überblick über die Daten gewinnen

- + Wie viele Landkreise haben wir in den Daten?
- + Wie ist die Verteilung der Schulden, Arbeitsenquote und des BIP?

Hierzu müssen wir erst noch die Arbeitslosenquote berechnen:

$$\text{Arbeitslosenquote} = \text{Erwerbslose} / (\text{Erwerbstätige} + \text{Erwerbslose})$$

```
# Zuerst wollen wir uns noch die Arbeitslosenquote pro Landkreis berechnen  
gesamtdaten <- gesamtdaten %>%  
  mutate(alo_quote = (total_alo / (erw+total_alo))*100)
```

Anzahl an Beobachtungen

Quick and dirty* (einfacher Tibble Datensatz): Einen Blick auf die Anzahl an Erwerbstätigen und Einwohnern in Deutschland werfen.

```
# Wie viele Erwerbstätige und Einwohner (ohne Berlin, Hamburg, Bremen und Bremerhaven) hat Deutschland?  
gesamtdaten %>%  
  summarise(total_erw = sum(erw, na.rm=TRUE), total_einwohner = sum(Einwohner, na.rm=TRUE))
```

```
## # A tibble: 1 x 2  
##   total_erw total_einwohner  
##       <dbl>          <dbl>  
## 1    41068448        76573483
```

Anzahl an Beobachtungen

Quick and dirty* (einfacher Tibble Datensatz): Einen Blick auf die Anzahl an Erwerbstätigen und Einwohnern in Deutschland werfen.

```
# Wie viele Erwerbstätige und Einwohner (ohne Berlin, Hamburg, Bremen und Bremerhaven) hat Deutschland?  
gesamtdaten %>%  
  summarise(total_erw = sum(erw, na.rm=TRUE), total_einwohner = sum(Einwohner, na.rm=TRUE))
```

```
## # A tibble: 1 x 2  
##   total_erw total_einwohner  
##       <dbl>          <dbl>  
## 1    41068448        76573483
```

- ✚ 41 Mio. Erwerbstätige und 76,5 Mio Einwohner in Deutschland
- ✚ Folgende Stadtstaaten sind nicht in unseren Berechnungen enthalten:
 - ✚ Hamburg (1.8 Mio.)
 - ✚ Berlin (3.75 Mio.)
 - ✚ Bremen (0.7 Mio.)
 - ✚ Bremerhaven (0.1 Mio.)

Anzahl an Beobachtungen

Etwas besser mit `skimr` Daten veranschaulichen

Anzahl an Beobachtungen

Etwas besser mit skimr Daten veranschaulichen

```
# Anschließend wollen wir eine Summary Statistic für alle Variablen ausgeben lassen
# Entfernen der Histogramme, damit alles auch schön in PDF gedruckt werden kann
gesamtdaten %>%
  select(alo_quote, Schulden_pro_kopf_lk, bip_pro_kopf, landkreis_name) %>%
  skim_without_charts() %>%
  summary()
```

Anzahl an Beobachtungen

Data summary

Name Piped data
Number of rows 401
Number of columns 4

Column type frequency:

factor 1
numeric 3

Group variables None

Anzahl an Beobachtungen

- ✚ 401 individuelle Beobachtungen in unserem Datensatz.

Hierbei handelt es sich um alle Landkreise und kreisfreien Städte in Deutschland.

| Stimmen diese Angaben?

Anzahl an Beobachtungen

- ✚ 401 individuelle Beobachtungen in unserem Datensatz.

Hierbei handelt es sich um alle Landkreise und kreisfreien Städte in Deutschland.

| Stimmen diese Angaben?

- ✚ In Deutschland gibt es 294 Landkreise.
- ✚ Weiterhin gibt es in Deutschland 107 kreisfreie Städte

(Quelle: Wikipedia)

Anzahl an Beobachtungen

Variable type: factor

skim_variable n_missing complete_rate ordered n_unique top_counts

landkreis_name 0 1 FALSE 379 Ans: 2, Asc: 2, Aug: 2, Bam: 2

- + Nur 379 unterschiedliche Landkreis Namen in unserem Datensatz mit 401 unterschiedlichen Beobachtungen (Regionalschlüsseln).

Woher kommt dies?

Anzahl an Beobachtungen

Variable type: factor

```
skim_variable  n_missing complete_rate ordered n_unique top_counts
```

	n_missing	complete_rate	ordered	n_unique	top_counts
landkreis_name	0	1 FALSE		379	Ans: 2, Asc: 2, Aug: 2, Bam: 2

- + Nur 379 unterschiedliche Landkreis Namen in unserem Datensatz mit 401 unterschiedlichen Beobachtungen (Regionalschlüsseln).

Woher kommt dies?

- + Stadt München ist eine Beobachtung
- + Landkreis München eine weitere Beobachtung

Beide haben unterschiedliche Regionalschlüssel. D.h. der "landkreis_name" ist der gleiche, jedoch ist der Regionalschlüssel ein anderer.

Anzahl an Beobachtungen

Nun möchten wir uns noch die einzelnen Variablen aus dem Datensatz näher anschauen:

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	5.33	2.36	1.66	3.38	4.99	6.88	13.52
Schulden_pro_kopf_lk	4	0.99	2742.91	2147.50	264.28	1295.03	2080.59	3447.76	14580.57
bip_pro_kopf	2	1.00	37086.95	16127.06	16398.46	27850.57	33105.40	40458.36	172436.71

Anzahl an Beobachtungen

- ✚ Fehlende Beobachtungen für Schulden pro Kopf: *vier* Landkreise
- ✚ Fehlende Beobachtungen für BIP pro Kopf: *zwei* Landkreise
- ✚ Fehlende Beobachtung für Einwohner: *vier* Landkreise

```
gesamtdaten %>%
  filter(is.na(Einwohner)) %>%
  select(landkreis_name)
```

```
## # A tibble: 4 x 1
##   landkreis_name
##   <fct>
## 1 Hamburg
## 2 Bremen
## 3 Bremerhaven
## 4 Berlin
```

Anzahl an Beobachtungen

- ✚ Fehlende Beobachtungen für Schulden pro Kopf: *vier* Landkreise
- ✚ Fehlende Beobachtungen für BIP pro Kopf: *zwei* Landkreise
- ✚ Fehlende Beobachtung für Einwohner: *vier* Landkreise

```
gesamtdaten %>%
  filter(is.na(Einwohner)) %>%
  select(landkreis_name)
```

```
## # A tibble: 4 x 1
##   landkreis_name
##   <fct>
## 1 Hamburg
## 2 Bremen
## 3 Bremerhaven
## 4 Berlin
```

Wir können diese Landkreise nicht mit in unsere Analyse mit einbeziehen auf Grund der fehlenden Informationen zu Einwohnern!

Beschreibung der Tabelle

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	5.33	2.36	1.66	3.38	4.99	6.88	13.52
Schulden_pro_kopf_lk	4	0.99	2742.91	2147.50	264.28	1295.03	2080.59	3447.76	14580.57
bip_pro_kopf	2	1.00	37086.95	16127.06	16398.46	27850.57	33105.40	40458.36	172436.71

Beschreibung der Tabelle

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	5.33	2.36	1.66	3.38	4.99	6.88	13.52
Schulden_pro_kopf_lk	4	0.99	2742.91	2147.50	264.28	1295.03	2080.59	3447.76	14580.57
bip_pro_kopf	2	1.00	37086.95	16127.06	16398.46	27850.57	33105.40	40458.36	172436.71

Bitte beschreiben Sie die Tabelle in ihren eigenen Worten!

Gehen Sie hierbei bitte auf eine Variable (alo_quote, Schulden_pro_Kopf_lk, bip_pro_kopf) und einen der folgenden Punkte ein:

- ✚ Mittelwert
- ✚ Standardabweichung
- ✚ Median

Beschreibung der Tabelle

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	5.33	2.36	1.66	3.38	4.99	6.88	13.52
Schulden_pro_kopf_lk	4	0.99	2742.91	2147.50	264.28	1295.03	2080.59	3447.76	14580.57
bip_pro_kopf	2	1.00	37086.95	16127.06	16398.46	27850.57	33105.40	40458.36	172436.71

Bitte beschreiben Sie die Tabelle in ihren eigenen Worten!

Gehen Sie hierbei bitte auf eine Variable (alo_quote, Schulden_pro_Kopf_lk, bip_pro_kopf) und einen der folgenden Punkte ein:

- ✚ Mittelwert
- ✚ Standardabweichung
- ✚ Median

05 : 00

Arbeitslosenquote

Mittelwert: 5,33 Prozent

- + Sehr hoch
- + Jedoch SGB II und SGB III
- + Konsistenzcheck auf [Statista](#) zeigt eine Arbeitslosenquote von 5,8% für 2017
- + **Jedoch:** Wir haben nicht Berlin und Hamburg in den Daten

Standardabweichung: 2,36

- + Sehr hohe Streuung
- + Deutliche regionale Unterschiede
- + Ist in Prozentpunkten

Median: 4,99 Prozent

- + Nahe am Mittelwert
- + Deutet darauf hin das es wenige Landkreise mit sehr extremen Ausreißern gibt

Verschuldung pro Kopf

Mittelwert: 2743€

- + Moderat von der Höhe her

Standardabweichung: 2148€

- + Sehr hohe Streuung
- + Deutliche regionale Unterschiede

Median: 2081€

- + Nahe am Mittelwert
- + Deutet darauf hin das es wenige Landkreise mit sehr extremen Ausreißern gibt

BIP pro Kopf

Mittelwert: 37086€

- + Insgesamt recht hoch
- + Starker Wirtschaftsstandort Deutschland

Standardabweichung: 16127€

- + Sehr hohe Streuung
- + Deutliche regionale Unterschiede
- + Könnte von einzelnen Landkreisen getrieben werden

Median: 33105€

- + Weiter weg vom Mittelwert
- + Deutet darauf hin das es einzelne Landkreise mit sehr extremen Ausreißern gibt

Die Arbeitslosenquote auf Bundeslandebene

Die Arbeitslosenquote auf Bundeslandebene

Es gibt deutliche Unterschiede in der Arbeitslosenquote über die Bundesländer hinweg!

Wir betrachten:

- ✚ Querschnittsdaten aus 2017
- ✚ Alle Landkreise
- ✚ Für einige Landkreise haben wir keine Informationen (sogenannte "Missing values" -> n_missing)

Was wollen wir?

Die regionale Verteilung der Arbeitslosenquote in Deutschland im Jahr 2017 näher betrachten.

Die Arbeitslosenquote auf Bundeslandebene

Zuerst aggregieren wir die Daten auf Bundeslandebene:

```
bula_data <- gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(alo_quote), median_alo = median(alo_quote)) %>%
  ungroup()
```

Die Arbeitslosenquote auf Bundeslandebene

count: false

gesamtdaten

```
## # A tibble: 401 x 12
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 7 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>
```

```
gesamtdaten %>%
```

```
  group_by( bundesland_name )
```

```
## # A tibble: 401 x 12
## # Groups:   bundesland_name [16]
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg   01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 7 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>
```

```
gesamtdaten %>%
```

```
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote),
            sd_alo = sd(alo_quote),
            median_alo = median(alo_quote))
```

```
## # A tibble: 16 x 4
##   bundesland_name     mean_alo    sd_alo median_alo
##   <fct>          <dbl>     <dbl>      <dbl>
## 1 Baden-Württemberg  3.31      0.642     3.33
## 2 Bayern             3.04      0.775     2.98
## 3 Berlin              NA        NA         NA
## 4 Brandenburg        7.98      1.93      8.27
## 5 Bremen              8.61      2.02      8.61
## 6 Hamburg             NA        NA         NA
## 7 Hessen              5.02      1.34      5.11
## 8 Mecklenburg-Vorpommern  8.41      1.56      7.95
## 9 Niedersachsen       6.14      1.76      6.03
## 10 Nordrhein-Westfalen 7.14      2.42      6.84
## 11 Rheinland-Pfalz    5.29      1.47      5.20
## 12 Saarland            5.88      1.75      5.34
## 13 Sachsen             6.59      1.05      6.25
## 14 Sachsen-Anhalt      9.00      1.56      8.87
## 15 Schleswig-Holstein  6.34      0.989     6.86
## 16 Thüringen           6.26      1.79      5.68
```

```
gesamtdaten %>%  
  group_by( bundesland_name ) %>%  
  summarise(mean_alo = mean(alo_quote),  
            sd_alo = sd(alo_quote),  
            median_alo = median(alo_quote)) %>%  
  ungroup() -> bula_data
```

Die Arbeitslosenquote auf Bundeslandebene

Anschließend wollen wir uns eine ansprechende und informative deskriptive Tabelle erstellen:

```
## # A tibble: 14 x 4
##   bundesland_name     mean_alo   sd_alo   median
##   <fct>          <dbl>    <dbl>    <
## 1 Bayern            3.04    0.775
## 2 Baden-Württemberg 3.31    0.642
## 3 Hessen             5.02    1.34
## 4 Rheinland-Pfalz   5.29    1.47
## 5 Saarland           5.88    1.75
## 6 Niedersachsen     6.14    1.76
## 7 Thüringen          6.26    1.79
## 8 Schleswig-Holstein 6.34    0.989
## 9 Sachsen             6.59    1.05
## 10 Nordrhein-Westfalen 7.14    2.42
## 11 Brandenburg        7.98    1.93
## 12 Mecklenburg-Vorpommern 8.41    1.56
## 13 Bremen              8.61    2.02
## 14 Sachsen-Anhalt      9.00    1.56
```

Die Arbeitslosenquote auf Bundeslandebene

Anschließend wollen wir uns eine ansprechende und informative deskriptive Tabelle erstellen:

```
## # A tibble: 14 x 4
##   bунdesland_name   mean_alo   sd_alo   median
##   <fct>          <dbl>    <dbl>    <dbl>
## 1 Bayern            3.04     0.775   3.04
## 2 Baden-Württemberg 3.31     0.642   3.31
## 3 Hessen            5.02     1.34    5.02
## 4 Rheinland-Pfalz  5.29     1.47    5.29
## 5 Saarland          5.88     1.75    5.88
## 6 Niedersachsen    6.14     1.76    6.14
## 7 Thüringen         6.26     1.79    6.26
## 8 Schleswig-Holstein 6.34    0.989   6.34
## 9 Sachsen           6.59     1.05    6.59
## 10 Nordrhein-Westfalen 7.14    2.42    7.14
## 11 Brandenburg      7.98     1.93    7.98
## 12 Mecklenburg-Vorpommern 8.41    1.56    8.41
## 13 Bremen            8.61     2.02    8.61
## 14 Sachsen-Anhalt   9.00     1.56    9.00
```

Bundesland	Arbeitslosenquote		
	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

Bitte beachten:

Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Tabelle aufgeführt wurden.

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Die Arbeitslosenquote auf Bundeslandebene

Die Darstellung mit dem Paket `kableExtra` ist deutlich ansprechender als nur einen Tibble zu zeigen!

Folgender Code wurde hier verwendet, welchen wir in der nächsten Folie Schritt für Schritt durchgehen werden:

```
bulu_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive")) %>%
  kable_paper(full_width = F) %>%
  row_spec(c(7, 9, 11, 12, 14), bold = T, color = "white", background = "#BBBBBB") %>%
  add_header_above(c(" " = 1, "Arbeitslosenquote" = 3), align = "c") %>%
  footnote(general = "Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Tabelle erscheinen.", general_title = "Bitte beachten:", number = "Die ostdeutschen Bundesländer sind grau hinterlegt.")
```

```
bula_data
```

```
## # A tibble: 16 x 4
##   bundesland_name     mean_alo   sd_alo median_alo
##   <fct>          <dbl>    <dbl>      <dbl>
## 1 Baden-Württemberg  3.31     0.642     3.33
## 2 Bayern            3.04     0.775     2.98
## 3 Berlin             NA       NA        NA
## 4 Brandenburg       7.98     1.93      8.27
## 5 Bremen             8.61     2.02      8.61
## 6 Hamburg            NA       NA        NA
## 7 Hessen              5.02     1.34      5.11
## 8 Mecklenburg-Vorpommern 8.41     1.56      7.95
## 9 Niedersachsen     6.14     1.76      6.03
## 10 Nordrhein-Westfalen 7.14     2.42      6.84
## 11 Rheinland-Pfalz  5.29     1.47      5.20
## 12 Saarland           5.88     1.75      5.34
## 13 Sachsen            6.59     1.05      6.25
## 14 Sachsen-Anhalt    9.00     1.56      8.87
## 15 Schleswig-Holstein 6.34     0.989     6.86
## 16 Thüringen          6.26     1.79      5.68
```

```
bula_data %>%
```

```
  arrange( mean_alo )
```

	bundesland_name	mean_alo	sd_alo	median_alo
	<fct>	<dbl>	<dbl>	<dbl>
## 1	Bayern	3.04	0.775	2.98
## 2	Baden-Württemberg	3.31	0.642	3.33
## 3	Hessen	5.02	1.34	5.11
## 4	Rheinland-Pfalz	5.29	1.47	5.20
## 5	Saarland	5.88	1.75	5.34
## 6	Niedersachsen	6.14	1.76	6.03
## 7	Thüringen	6.26	1.79	5.68
## 8	Schleswig-Holstein	6.34	0.989	6.86
## 9	Sachsen	6.59	1.05	6.25
## 10	Nordrhein-Westfalen	7.14	2.42	6.84
## 11	Brandenburg	7.98	1.93	8.27
## 12	Mecklenburg-Vorpommern	8.41	1.56	7.95
## 13	Bremen	8.61	2.02	8.61
## 14	Sachsen-Anhalt	9.00	1.56	8.87
## 15	Berlin	NA	NA	NA
## 16	Hamburg	NA	NA	NA

```
bula_data %>%  
  arrange( mean_alo ) %>%  
  filter( !is.na(mean_alo) )
```

```
## # A tibble: 14 x 4  
##   bundesland_name      mean_alo    sd_alo median_alo  
##   <fct>          <dbl>     <dbl>      <dbl>  
## 1 Bayern            3.04     0.775     2.98  
## 2 Baden-Württemberg 3.31     0.642     3.33  
## 3 Hessen             5.02     1.34      5.11  
## 4 Rheinland-Pfalz   5.29     1.47      5.20  
## 5 Saarland           5.88     1.75      5.34  
## 6 Niedersachsen     6.14     1.76      6.03  
## 7 Thüringen          6.26     1.79      5.68  
## 8 Schleswig-Holstein 6.34     0.989     6.86  
## 9 Sachsen            6.59     1.05      6.25  
## 10 Nordrhein-Westfalen 7.14     2.42      6.84  
## 11 Brandenburg       7.98     1.93      8.27  
## 12 Mecklenburg-Vorpommern 8.41     1.56      7.95  
## 13 Bremen             8.61     2.02      8.61  
## 14 Sachsen-Anhalt    9.00     1.56      8.87
```

```
bula_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2)
```

Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

```

bula_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "ho

```

Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

```

bula_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
  kable_paper(full_width = F)

```

Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

```

bula_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "ho
  kable_paper(full_width = F) %>%
  row_spec(c(7,9, 11,12,14), bold = T, color = "whit

```

Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

```

bula_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "ho
  kable_paper(full_width = F) %>%
  row_spec(c(7,9, 11,12,14), bold = T, color = "whit
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Arbeitslosenquote				
Bundesland	Mittelwert	Std.	Median	
Bayern	3.04	0.77	2.98	
Baden-Württemberg	3.31	0.64	3.33	
Hessen	5.02	1.34	5.11	
Rheinland-Pfalz	5.29	1.47	5.20	
Saarland	5.88	1.75	5.34	
Niedersachsen	6.14	1.76	6.03	
Thüringen	6.26	1.79	5.68	
Schleswig-Holstein	6.34	0.99	6.86	
Sachsen	6.59	1.05	6.25	
Nordrhein-Westfalen	7.14	2.42	6.84	
Brandenburg	7.98	1.93	8.27	
Mecklenburg-Vorpommern	8.41	1.56	7.95	
Bremen	8.61	2.02	8.61	
Sachsen-Anhalt	9.00	1.56	8.87	

```

bula_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "ho
  kable_paper(full_width = F) %>%
  row_spec(c(7,9, 11,12,14), bold = T, color = "whit
  add_header_above(c(" " = 1, "Arbeitslosenquote" =
  footnote(general = "Wir haben keine Informationen
  general_title = "Bitte beachten: ",
  number = "Die ostdeutschen Bundesländer s

```

Arbeitslosenquote			
Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

Bitte beachten:

Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Tabelle aufgeführt wurden.

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Arbeitslosenquote

Bundesland	Mittelwert	Std.	Median
Bayern	3.04	0.77	2.98
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.02	1.34	5.11
Rheinland-Pfalz	5.29	1.47	5.20
Saarland	5.88	1.75	5.34
Niedersachsen	6.14	1.76	6.03
Thüringen	6.26	1.79	5.68
Schleswig-Holstein	6.34	0.99	6.86
Sachsen	6.59	1.05	6.25
Nordrhein-Westfalen	7.14	2.42	6.84
Brandenburg	7.98	1.93	8.27
Mecklenburg-Vorpommern	8.41	1.56	7.95
Bremen	8.61	2.02	8.61
Sachsen-Anhalt	9.00	1.56	8.87

Bitte beachten:

Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Tabelle aufgeführt wurden.

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Länder im Süden Deutschlands haben eine sehr niedrige Arbeitslosenquote (<4%)

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Länder im Süden Deutschlands haben eine sehr niedrige Arbeitslosenquote (<4%)
- + Ostdeutsche Bundesländer leider unter hohen Arbeitslosenquoten (>8%)

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Länder im Süden Deutschlands haben eine sehr niedrige Arbeitslosenquote (<4%)
- + Ostdeutsche Bundesländer leider unter hohen Arbeitslosenquoten (>8%)
- + Standardabweichung bei allen Bundesländern vergleichbar
 - + Es gibt hier vermutlich keine großen Ausreißer bei den Landkreisen

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Länder im Süden Deutschlands haben eine sehr niedrige Arbeitslosenquote (<4%)
- + Ostdeutsche Bundesländer leider unter hohen Arbeitslosenquoten (>8%)
- + Standardabweichung bei allen Bundesländern vergleichbar
 - + Es gibt hier vermutlich keine großen Ausreißer bei den Landkreisen
- + Median liegt recht nahe am Mittelwert für die Bundesländern

Allgemein sehen wir hier sehr große Unterschiede zwischen Ost- und Westdeutschland!

Die Arbeitslosenquote zwischen Ost- und Westdeutschland

Wir wollen uns eine neue Variable "ost", bzw. "ost_name" generieren. Anschließend können wir uns die Arbeitslosigkeit für Ost- und Westdeutschland anschauen.

```
gesamtdaten <- gesamtdaten %>%
  mutate( ost = as.factor(ifelse(bundesland_name %in% c("Brandenburg", "Mecklenburg-Vorpommern", "Sachsen", "Sachsen-Anhalt", "Thüringen"),
  ost_name = ifelse(ost == 1, "Ostdeutschland", "Westdeutschland")))
```

gesamtdaten

```
## # A tibble: 401 x 12
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 7 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>
```

```
gesamtdaten %>%
```

```
  mutate( ost = as.factor(ifelse(bundesland_name %in%
    ost_name = ifelse(ost == 1, "Ostdeutschlan
```

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>          <dbl> <fct>      <chr>      <fct>
## 1 01001           4512 Flensburg 01 Schleswig-Hols...
## 2 01002           12345 Kiel     01 Schleswig-Hols...
## 3 01003           9692 Lübeck   01 Schleswig-Hols...
## 4 01004           3836 Neumünster 01 Schleswig-Hols...
## 5 01051           4632 Dithmarschen 01 Schleswig-Hols...
## 6 01053           5592 Herzogtum Lau... 01 Schleswig-Hols...
## 7 01054           5657 Nordfriesland 01 Schleswig-Hols...
## 8 01055           5748 Ostholstein 01 Schleswig-Hols...
## 9 01056           8599 Pinneberg 01 Schleswig-Hols...
## 10 01057          3264 Plön     01 Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

Die Arbeitslosenquote zwischen Ost- und Westdeutschland

```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), sd_alo = sd(alo_quote, na.rm = T), min_alo = min(alo_quote,
ungroup() %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Minimum",
                    "P25",
                    "Median",
                    "P75",
                    "Maximum"), digits = 2) %>%
kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive")) %>%
kable_paper(full_width = F) %>%
add_header_above(c(" " = 1, "Arbeitslosenquote" = 7), align = "c") %>%
footnote(general = "Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Berechnur
general_title = "Bitte beachten: ")
```

gesamtdaten

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck        01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen  01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland  01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
  group_by(ost_name)
```

```
## # A tibble: 401 x 14
## # Groups:   ost_name [2]
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg   01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), s
```



```
## # A tibble: 2 x 8
##   ost_name      mean_alo    sd_alo  min_alo    q25 median_alo    q75 max_alo
##   <chr>        <dbl>     <dbl>    <dbl>    <dbl>     <dbl>     <dbl>    <dbl>
## 1 Ostdeutschland 7.46     1.95    3.76    6.00    7.30    8.85    12.9
## 2 Westdeutschland 4.83     2.17    1.66    3.23    4.27    6.17    13.5
```

```
gesamtdaten %>%  
  group_by(ost_name) %>%  
  summarise(mean_alo = mean(alo_quote, na.rm = T), s  
ungroup()
```

```
## # A tibble: 2 x 8  
##   ost_name      mean_alo    sd_alo  min_alo    q25 median_alo    q75 max_alo  
##   <chr>        <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 Ostdeutschland    7.46    1.95    3.76    6.00    7.30    8.85    12.9  
## 2 Westdeutschland   4.83    2.17    1.66    3.23    4.27    6.17    13.5
```

```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), s
ungroup() %>%
  kbl(col.names = c("Bundesland",
                  "Mittelwert",
                  "Std.",
                  "Minimum",
                  "P25",
                  "Median",
                  "P75",
                  "Maximum"), digits = 2)
```

Bundesland	Mittelwert	Std.	Minimum	P25	Median	P75	Maximum
Ostdeutschland	7.46	1.95	3.76	6.00	7.30	8.85	12.85
Westdeutschland	4.83	2.17	1.66	3.23	4.27	6.17	13.52

```

gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), s
ungroup() %>%
  kbl(col.names = c("Bundesland",
                  "Mittelwert",
                  "Std.",
                  "Minimum",
                  "P25",
                  "Median",
                  "P75",
                  "Maximum"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "ho

```

Bundesland	Mittelwert	Std.	Minimum	P25	Median	P75	Maximum
Ostdeutschland	7.46	1.95	3.76	6.00	7.30	8.85	12.85
Westdeutschland	4.83	2.17	1.66	3.23	4.27	6.17	13.52

```

gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), s
ungroup() %>%
  kbl(col.names = c("Bundesland",
                  "Mittelwert",
                  "Std.",
                  "Minimum",
                  "P25",
                  "Median",
                  "P75",
                  "Maximum"), digits = 2) %>%
kable_styling(bootstrap_options = c("striped", "ho
kable_paper(full_width = F)

```

Bundesland	Mittelwert	Std.	Minimum	P25	Median	P75	Maximum
Ostdeutschland	7.46	1.95	3.76	6.00	7.30	8.85	12.85
Westdeutschland	4.83	2.17	1.66	3.23	4.27	6.17	13.52

```

gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), s
ungroup() %>%
  kbl(col.names = c("Bundesland",
                  "Mittelwert",
                  "Std.",
                  "Minimum",
                  "P25",
                  "Median",
                  "P75",
                  "Maximum"), digits = 2) %>%
kable_styling(bootstrap_options = c("striped", "ho
kable_paper(full_width = F) %>%
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Arbeitslosenquote

Bundesland	Mittelwert	Std.	Minimum	P25	Median	P75	Maximum
Ostdeutschland	7.46	1.95	3.76	6.00	7.30	8.85	12.85
Westdeutschland	4.83	2.17	1.66	3.23	4.27	6.17	13.52

```

gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(mean_alo = mean(alo_quote, na.rm = T), s
ungroup() %>%
  kbl(col.names = c("Bundesland",
                  "Mittelwert",
                  "Std.",
                  "Minimum",
                  "P25",
                  "Median",
                  "P75",
                  "Maximum"), digits = 2) %>%
kable_styling(bootstrap_options = c("striped", "ho
kable_paper(full_width = F) %>%
add_header_above(c(" " = 1, "Arbeitslosenquote" =
footnote(general = "Wir haben keine Informationen
      general_title = "Bitte beachten: ")

```

Arbeitslosenquote

Bundesland	Mittelwert	Std.	Minimum	P25	Median	P75	Maximum
Ostdeutschland	7.46	1.95	3.76	6.00	7.30	8.85	12.85
Westdeutschland	4.83	2.17	1.66	3.23	4.27	6.17	13.52

Bitte beachten:

Wir haben keine Informationen zu Berlin und Hamburg, weshalb sie nicht in der Berechnung enthalten sind.

Die Arbeitslosenquote zwischen Ost- und Westdeutschland

Große Unterschiede werden sichtbar:

- ✚ Fast 3 Prozentpunkte niedriger in den westdeutschen Bundesländern
- ✚ Die Standardabweichung ist vergleichbar
- ✚ Der Median liegt in Westdeutschland sogar noch deutlicher unter dem Mittelwert als in Ostdeutschland
- ✚ Im **25% Quantil in Ostdeutschland** ist die Arbeitslosenquote bei **6,00%**
- ✚ In **Westdeutschland** ist das **75% Quantil** bei einer Arbeitslosenquote von **6,17%**!

Arbeitslosenquote, BIP pro Kopf und Schulden pro Kopf

```
gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(alo_quote), mean_bip_kopf = mean(bip_pro_kopf), sd_bip_ko
ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.", "Mittelwert", "Std.", "Mittelwert", "Std."), digits = 2,
    caption = "Deskriptive Tabelle komplett") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive")) %>%
  kable_paper(full_width = F) %>%
  row_spec(c(7,9, 11,12,13), bold = T, color = "white", background = "#BBBBBB") %>%
  add_header_above(c(" " = 1, "Arbeitslosenquote" = 2, "BIP pro Kopf" = 2, "Schulden pro Kopf" = 2), align =
footnote(general = "Wir haben keine Informationen zu Berlin, Hamburg und Bremen bzgl. ihrer Schulden pro Ko
  general_title = "Bitte beachten: ",
  number = "Die ostdeutschen Bundesländer sind grau hinterlegt.")
```

gesamtdaten

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  group_by( bundesland_name )
```

```
## # A tibble: 401 x 14
## # Groups:   bundesland_name [16]
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
```

```
## # A tibble: 16 x 7
##   bundesland_name  mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_k...
##   <fct>          <dbl>    <dbl>      <dbl>       <dbl>       <dbl>
## 1 Baden-Württemb...  3.31    0.642     43658.     12454.     2185.
## 2 Bayern           3.04    0.775     43002.     19977.     1897.
## 3 Berlin            NA     NA        NA         NA         NA
## 4 Brandenburg      7.98    1.93      28971.     5817.      2831.
## 5 Bremen            8.61    2.02      43798.     10143.     NA
## 6 Hamburg           NA     NA        NA         NA         NA
## 7 Hessen            5.02    1.34      40203.     16477.     3731.
## 8 Mecklenburg-Vo...  8.41    1.56      27908.     6385.      3565.
## 9 Niedersachsen    6.14    1.76      35514.     23189.     1941.
## 10 Nordrhein-West... 7.14    2.42      36165.     10883.     4243.
## 11 Rheinland-Pfalz  5.29    1.47      34918.     14184.     3131.
## 12 Saarland          5.88    1.75      33075.     7271.      5959.
## 13 Sachsen           6.59    1.05      28731.     5083.      2306.
## 14 Sachsen-Anhalt   9.00    1.56      27218.     3853.      2809.
## 15 Schleswig-Hols...  6.34    0.989     32777.     7412.      2602.
## 16 Thüringen         6.26    1.79      28452.     5381.      2832.
## # ... with 1 more variable: sd_schulden <dbl>
```

```
gesamtdataen %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
    ungroup() -> bula_data_all
```

```
gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all
```

```
## # A tibble: 16 x 7
##   bundesland_name mean_alo  sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_k...
##   <fct>          <dbl>    <dbl>      <dbl>       <dbl>       <dbl>
## 1 Baden-Württemb...     3.31    0.642     43658.     12454.     2185.
## 2 Bayern            3.04    0.775     43002.     19977.     1897.
## 3 Berlin             NA      NA        NA         NA         NA
## 4 Brandenburg        7.98    1.93      28971.     5817.      2831.
## 5 Bremen              8.61    2.02      43798.     10143.     NA
## 6 Hamburg             NA      NA        NA         NA         NA
## 7 Hessen              5.02    1.34      40203.     16477.     3731.
## 8 Mecklenburg-Vo...   8.41    1.56      27908.     6385.      3565.
## 9 Niedersachsen       6.14    1.76      35514.     23189.     1941.
## 10 Nordrhein-West...  7.14    2.42      36165.     10883.     4243.
## 11 Rheinland-Pfalz   5.29    1.47      34918.     14184.     3131.
## 12 Saarland            5.88   1.75      33075.     7271.      5959.
## 13 Sachsen             6.59   1.05      28731.     5083.      2306.
## 14 Sachsen-Anhalt     9.00    1.56      27218.     3853.      2809.
## 15 Schleswig-Hols...   6.34    0.989     32777.     7412.      2602.
## 16 Thüringen           6.26   1.79      28452.     5381.      2832.
## # ... with 1 more variable: sd_schulden <dbl>
```

```
gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo )
```

```
## # A tibble: 16 x 7
##   bundesland_name mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_k...
##   <fct>          <dbl>    <dbl>      <dbl>       <dbl>       <dbl>
## 1 Bayern           3.04    0.775     43002.     19977.     1897.
## 2 Baden-Württemb... 3.31    0.642     43658.     12454.     2185.
## 3 Hessen            5.02   1.34      40203.     16477.     3731.
## 4 Rheinland-Pfalz  5.29    1.47      34918.     14184.     3131.
## 5 Saarland          5.88   1.75      33075.     7271.      5959.
## 6 Niedersachsen    6.14    1.76      35514.     23189.     1941.
## 7 Thüringen         6.26    1.79      28452.     5381.      2832.
## 8 Schleswig-Hols...  6.34    0.989     32777.     7412.      2602.
## 9 Sachsen           6.59    1.05      28731.     5083.      2306.
## 10 Nordrhein-West... 7.14    2.42      36165.     10883.     4243.
## 11 Brandenburg      7.98    1.93      28971.     5817.      2831.
## 12 Mecklenburg-Vo...  8.41    1.56      27908.     6385.      3565.
## 13 Bremen            8.61    2.02      43798.     10143.     NA
## 14 Sachsen-Anhalt   9.00    1.56      27218.     3853.      2809.
## 15 Berlin             NA     NA          NA        NA        NA
## 16 Hamburg            NA     NA          NA        NA        NA
## # ... with 1 more variable: sd_schulden <dbl>
```

```
gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) )
```

```
## # A tibble: 13 x 7
##   bundesland_name mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_k...
##   <fct>           <dbl>    <dbl>      <dbl>       <dbl>       <dbl>
## 1 Bayern            3.04    0.775     43002.     19977.     1897.
## 2 Baden-Württemb...  3.31    0.642     43658.     12454.     2185.
## 3 Hessen             5.02    1.34      40203.     16477.     3731.
## 4 Rheinland-Pfalz  5.29    1.47      34918.     14184.     3131.
## 5 Saarland           5.88    1.75      33075.     7271.      5959.
## 6 Niedersachsen     6.14    1.76      35514.     23189.     1941.
## 7 Thüringen          6.26    1.79      28452.     5381.      2832.
## 8 Schleswig-Hols...  6.34    0.989     32777.     7412.      2602.
## 9 Sachsen             6.59    1.05      28731.     5083.      2306.
## 10 Nordrhein-West... 7.14    2.42      36165.     10883.     4243.
## 11 Brandenburg        7.98    1.93      28971.     5817.      2831.
## 12 Mecklenburg-Vo...  8.41    1.56      27908.     6385.      3565.
## 13 Sachsen-Anhalt    9.00    1.56      27218.     3853.      2809.
## # ... with 1 more variable: sd_schulden <dbl>
```

```

gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.
  caption = "Deskriptive Tabelle komplett")

```

Bundesland	Deskriptive Tabelle komplett				
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert
Bayern	3.04	0.77	4300	1.64	19976.60
Baden-Württemberg	3.31	0.64	4365	8.30	12454.30
Hessen	5.02	1.34	4020	3.08	16476.82
Rheinland-Pfalz	5.29	1.47	3491	7.53	14184.07
Saarland	5.88	1.75	3307	5.37	7270.71
Niedersachsen	6.14	1.76	3551	4.14	23188.65
Thüringen	6.26	1.79	2845	1.50	5380.95
Schleswig-Holstein	6.34	0.99	3277	7.38	7411.63
Sachsen	6.59	1.05	2873	0.71	5082.77
Nordrhein-Westfalen	7.14	2.42	3616	4.76	10883.06
Brandenburg	7.98	1.93	2897	0.85	5816.92
Mecklenburg-Vorpommern	8.41	1.56	2790	8.14	6384.54
Sachsen-Anhalt	9.00	1.56	2721	8.33	3852.69

```

gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.
    caption = "Deskriptive Tabelle komplett") %>%
  kable_styling(bootstrap_options = c("striped", "ho
                font_size = 9)

```

Deskriptive Tabelle komplett

Bundesland	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.04	0.77	43001.64	19976.60	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43658.30	12454.30	2185.49	1612.86
Hessen	5.02	1.34	40203.08	16476.82	3730.85	2870.38
Rheinland-Pfalz	5.29	1.47	34917.53	14184.07	3130.70	3593.36
Saarland	5.88	1.75	33075.37	7270.71	5958.80	1363.06
Niedersachsen	6.14	1.76	35514.14	23188.65	1941.34	1381.10
Thüringen	6.26	1.79	28451.50	5380.95	2832.09	547.15
Schleswig-Holstein	6.34	0.99	32777.38	7411.63	2601.65	1700.98
Sachsen	6.59	1.05	28730.71	5082.77	2305.58	674.06
Nordrhein-Westfalen	7.14	2.42	36164.76	10883.06	4243.47	2483.34
Brandenburg	7.98	1.93	28970.85	5816.92	2831.16	1473.58
Mecklenburg-Vorpommern	8.41	1.56	27908.14	6384.54	3564.62	1667.49
Sachsen-Anhalt	9.00	1.56	27218.33	3852.69	2809.25	1092.19

```

gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.
    caption = "Deskriptive Tabelle komplett") %>%
  kable_styling(bootstrap_options = c("striped", "ho
      font_size = 9) %>%
  kable_paper(full_width = F)

```

Deskriptive Tabelle komplett

Bundesland	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.04	0.77	43001.64	19976.60	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43658.30	12454.30	2185.49	1612.86
Hessen	5.02	1.34	40203.08	16476.82	3730.85	2870.38
Rheinland-Pfalz	5.29	1.47	34917.53	14184.07	3130.70	3593.36
Saarland	5.88	1.75	33075.37	7270.71	5958.80	1363.06
Niedersachsen	6.14	1.76	35514.14	23188.65	1941.34	1381.10
Thüringen	6.26	1.79	28451.50	5380.95	2832.09	547.15
Schleswig-Holstein	6.34	0.99	32777.38	7411.63	2601.65	1700.98
Sachsen	6.59	1.05	28730.71	5082.77	2305.58	674.06
Nordrhein-Westfalen	7.14	2.42	36164.76	10883.06	4243.47	2483.34
Brandenburg	7.98	1.93	28970.85	5816.92	2831.16	1473.58
Mecklenburg-Vorpommern	8.41	1.56	27908.14	6384.54	3564.62	1667.49
Sachsen-Anhalt	9.00	1.56	27218.33	3852.69	2809.25	1092.19

```

gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.
  caption = "Deskriptive Tabelle komplett") %>%
kable_styling(bootstrap_options = c("striped", "ho
          font_size = 9) %>%
  kable_paper(full_width = F) %>%
  row_spec(c(7,9, 11,12,13), bold = T, color = "whit

```

Deskriptive Tabelle komplett

Bundesland	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.04	0.77	43001.64	19976.60	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43658.30	12454.30	2185.49	1612.86
Hessen	5.02	1.34	40203.08	16476.82	3730.85	2870.38
Rheinland-Pfalz	5.29	1.47	34917.53	14184.07	3130.70	3593.36
Saarland	5.88	1.75	33075.37	7270.71	5958.80	1363.06
Niedersachsen	6.14	1.76	35514.14	23188.65	1941.34	1381.10
Thüringen	6.26	1.79	28451.50	5380.95	2832.09	547.15
Schleswig-Holstein	6.34	0.99	32777.38	7411.63	2601.65	1700.98
Sachsen	6.59	1.05	28730.71	5082.77	2305.58	674.06
Nordrhein-Westfalen	7.14	2.42	36164.76	10883.06	4243.47	2483.34
Brandenburg	7.98	1.93	28970.85	5816.92	2831.16	1473.58
Mecklenburg-Vorpommern	8.41	1.56	27908.14	6384.54	3564.62	1667.49
Sachsen-Anhalt	9.00	1.56	27218.33	3852.69	2809.25	1092.19

```

gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.
    caption = "Deskriptive Tabelle komplett") %>%
  kable_styling(bootstrap_options = c("striped", "ho
      font_size = 9) %>%
  kable_paper(full_width = F) %>%
  row_spec(c(7,9, 11,12,13), bold = T, color = "whit
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Deskriptive Tabelle komplett

Bundesland	Arbeitslosenquote		BIP pro Kopf		Schulden pro Kopf	
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.04	0.77	43001.64	19976.60	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43658.30	12454.30	2185.49	1612.86
Hessen	5.02	1.34	40203.08	16476.82	3730.85	2870.38
Rheinland-Pfalz	5.29	1.47	34917.53	14184.07	3130.70	3593.36
Saarland	5.88	1.75	33075.37	7270.71	5958.80	1363.06
Niedersachsen	6.14	1.76	35514.14	23188.65	1941.34	1381.10
Thüringen	6.26	1.79	28451.50	5380.95	2832.09	547.15
Schleswig-Holstein	6.34	0.99	32777.38	7411.63	2601.65	1700.98
Sachsen	6.59	1.05	28730.71	5082.77	2305.58	674.06
Nordrhein-Westfalen	7.14	2.42	36164.76	10883.06	4243.47	2483.34
Brandenburg	7.98	1.93	28970.85	5816.92	2831.16	1473.58
Mecklenburg-Vorpommern	8.41	1.56	27908.14	6384.54	3564.62	1667.49
Sachsen-Anhalt	9.00	1.56	27218.33	3852.69	2809.25	1092.19

```

gesamtdaten %>%
  group_by( bundesland_name ) %>%
  summarise(mean_alo = mean(alo_quote), sd_alo = sd(
  ungroup() -> bula_data_all

bula_data_all %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_schulden_kopf) ) %>%
  kbl(col.names = c("Bundesland", "Mittelwert", "Std.
    caption = "Deskriptive Tabelle komplett") %>%
  kable_styling(bootstrap_options = c("striped", "ho
      font_size = 9) %>%
  kable_paper(full_width = F) %>%
  row_spec(c(7, 9, 11, 12, 13), bold = T, color = "whit
  add_header_above(c(" " = 1, "Arbeitslosenquote" =
  footnote(general = "Wir haben keine Informationen
      general_title = "Bitte beachten: ",
      number = "Die ostdeutschen Bundesländer s

```

Deskriptive Tabelle komplett

Bundesland	Arbeitslosenquote		BIP pro Kopf		Schulden pro Kopf	
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.04	0.77	43001.64	19976.60	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43658.30	12454.30	2185.49	1612.86
Hessen	5.02	1.34	40203.08	16476.82	3730.85	2870.38
Rheinland-Pfalz	5.29	1.47	34917.53	14184.07	3130.70	3593.36
Saarland	5.88	1.75	33075.37	7270.71	5958.80	1363.06
Niedersachsen	6.14	1.76	35514.14	23188.65	1941.34	1381.10
Thüringen	6.26	1.79	28451.50	5380.95	2832.09	547.15
Schleswig-Holstein	6.34	0.99	32777.38	7411.63	2601.65	1700.98
Sachsen	6.59	1.05	28730.71	5082.77	2305.58	674.06
Nordrhein-Westfalen	7.14	2.42	36164.76	10883.06	4243.47	2483.34
Brandenburg	7.98	1.93	28970.85	5816.92	2831.16	1473.58
Mecklenburg-Vorpommern	8.41	1.56	27908.14	6384.54	3564.62	1667.49
Sachsen-Anhalt	9.00	1.56	27218.33	3852.69	2809.25	1092.19

Bitte beachten:

Wir haben keine Informationen zu Berlin, Hamburg und Bremen bzgl. ihrer Schulden pro Kopf, weshalb sie nicht in der Tabelle aufgeführt wurden.

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Arbeitslosenquote, BIP pro Kopf und Schulden pro Kopf

Deskriptive Tabelle komplett

Bundesland	Arbeitslosenquote		BIP pro Kopf		Schulden pro Kopf	
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.04	0.77	43001.64	19976.60	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43658.30	12454.30	2185.49	1612.86
Hessen	5.02	1.34	40203.08	16476.82	3730.85	2870.38
Rheinland-Pfalz	5.29	1.47	34917.53	14184.07	3130.70	3593.36
Saarland	5.88	1.75	33075.37	7270.71	5958.80	1363.06
Niedersachsen	6.14	1.76	35514.14	23188.65	1941.34	1381.10
Thüringen	6.26	1.79	28451.50	5380.95	2832.09	547.15
Schleswig-Holstein	6.34	0.99	32777.38	7411.63	2601.65	1700.98
Sachsen	6.59	1.05	28730.71	5082.77	2305.58	674.06
Nordrhein-Westfalen	7.14	2.42	36164.76	10883.06	4243.47	2483.34
Brandenburg	7.98	1.93	28970.85	5816.92	2831.16	1473.58
Mecklenburg-Vorpommern	8.41	1.56	27908.14	6384.54	3564.62	1667.49
Sachsen-Anhalt	9.00	1.56	27218.33	3852.69	2809.25	1092.19

Bitte beachten:

Wir haben keine Informationen zu Berlin, Hamburg und Bremen bzgl. ihrer Schulden pro Kopf, weshalb sie nicht in der Tabelle aufgeführt wurden.

Arbeitslosenquote, BIP pro Kopf und Schulden pro Kopf

- ✚ Bundesländer mit niedrigen Arbeitslosenquoten haben ein hohes BIP pro Kopf
- ✚ Ostdeutsche Bundesländer haben alle ein BIP pro Kopf $< 30000\text{€}$
- ✚ Westdeutsche Bundesländer haben alle ein BIP pro Kopf $> 30000\text{€}$
- ✚ Einige Bundesländer mit wenig, einige mit hohen Schulden pro Kopf

Arbeitslosenquote, BIP pro Kopf und Schulden pro Kopf

- ✚ Bundesländer mit niedrigen Arbeitslosenquoten haben ein hohes BIP pro Kopf
- ✚ Ostdeutsche Bundesländer haben alle ein BIP pro Kopf $< 30000\text{€}$
- ✚ Westdeutsche Bundesländer haben alle ein BIP pro Kopf $> 30000\text{€}$
- ✚ Einige Bundesländer mit wenig, einige mit hohen Schulden pro Kopf

Allein durch Mittelwert und Standardabweichung können wir bereits sehr viel über regionale Unterschiede lernen.

Entwicklung des BIP

Auch zeitliche Entwicklungen können in einer Tabelle dargestellt werden

Als Beispiel sollten Sie sich die Tabelle zur Entwicklung des BIP pro Kopf in der Case-Study anschauen

Datenvisualisierung

Arbeitslosenquote

Das Auge verarbeitet Informationen deutlich schneller und intuitiver wenn diese in einer Grafik präsentiert werden, anstatt in Tabellenform.

Arbeitslosenquote

Das Auge verarbeitet Informationen deutlich schneller und intuitiver wenn diese in einer Grafik präsentiert werden, anstatt in Tabellenform.

Daher ist es wichtig Grafiken in den deskriptiven Analysen mit einzubeziehen

Daten: Querschnittsdaten zur Arbeitslosigkeit in den Landkreisen aus dem Jahr 2017

Arbeitslosenquote

Das Auge verarbeitet Informationen deutlich schneller und intuitiver wenn diese in einer Grafik präsentiert werden, anstatt in Tabellenform.

Daher ist es wichtig Grafiken in den deskriptiven Analysen mit einzubeziehen

Daten: Querschnittsdaten zur Arbeitslosigkeit in den Landkreisen aus dem Jahr 2017

Die folgende Grafik sollte enthalten:

- ✚ **Zeige alle Daten:** Jeder Landkreis wird durch einen Punkt in der Grafik repräsentiert
- ✚ Boxplot der Arbeitslosigkeit wird über die Punktwolke gelegt

gesamtdaten

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  select(alo_quote, landkreis_name, bundesland_name,
```

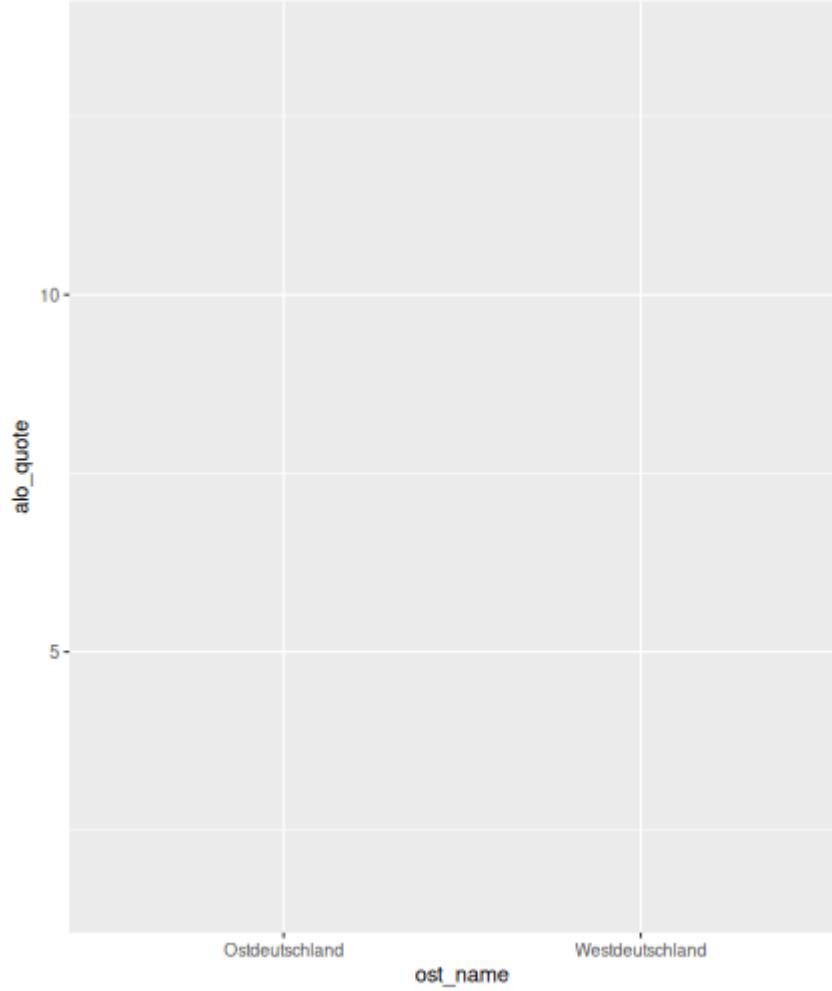
```
## # A tibble: 401 x 4
##   alo_quote landkreis_name     bundesland_name    ost_name
##       <dbl> <fct>           <fct>             <chr>
## 1     6.99 Flensburg        Schleswig-Holstein Westdeutschland
## 2     6.64 Kiel              Schleswig-Holstein Westdeutschland
## 3     7.02 Lübeck            Schleswig-Holstein Westdeutschland
## 4     6.90 Neumünster       Schleswig-Holstein Westdeutschland
## 5     7.09 Dithmarschen     Schleswig-Holstein Westdeutschland
## 6     7.36 Herzogtum Lauenburg Schleswig-Holstein Westdeutschland
## 7     5.86 Nordfriesland     Schleswig-Holstein Westdeutschland
## 8     5.90 Ostholstein       Schleswig-Holstein Westdeutschland
## 9     6.16 Pinneberg         Schleswig-Holstein Westdeutschland
## 10    6.90 Plön              Schleswig-Holstein Westdeutschland
## # ... with 391 more rows
```

```
gesamtdaten %>%
```

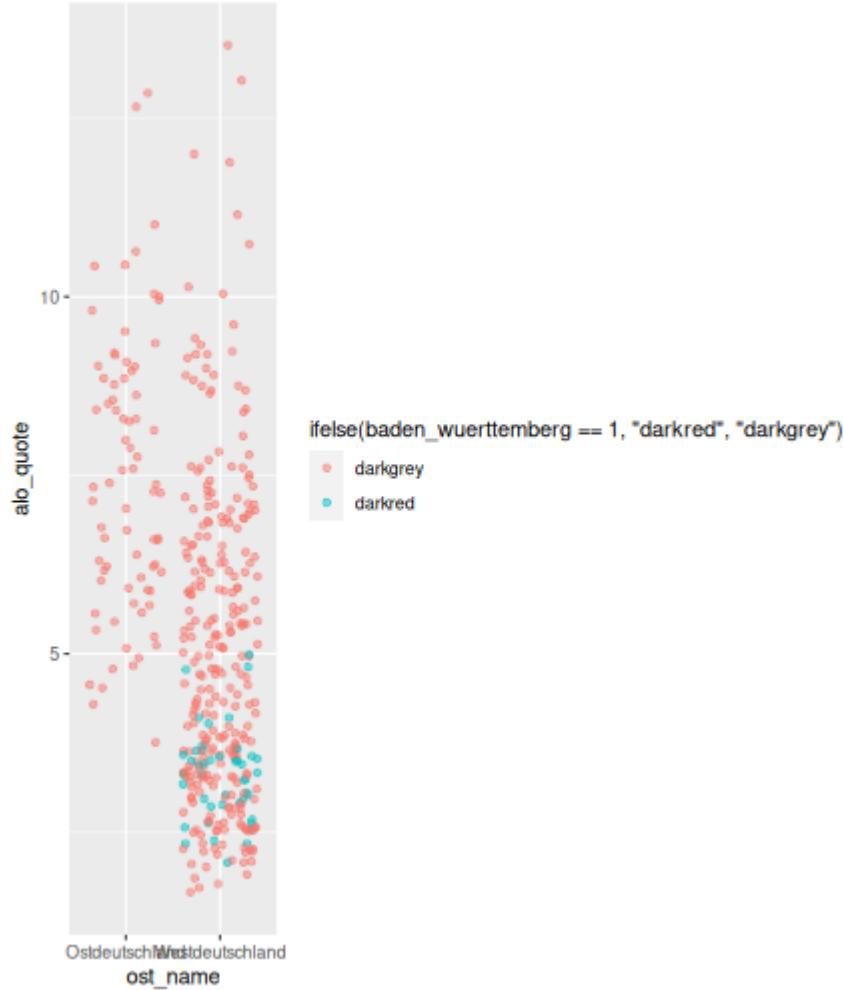
```
  select(alo_quote, landkreis_name, bundesland_name,  
         mutate(baden_wuerttemberg = as.factor(ifelse(bunde
```

```
## # A tibble: 401 x 5  
##   alo_quote landkreis_name  bundesland_name ost_name baden_wuerttemberg  
##       <dbl> <fct>          <fct>           <chr>    <fct>  
## 1 6.99 Flensburg Schleswig-Holste... Westdeutsch... 0  
## 2 6.64 Kiel      Schleswig-Holste... Westdeutsch... 0  
## 3 7.02 Lübeck   Schleswig-Holste... Westdeutsch... 0  
## 4 6.90 Neumünster Schleswig-Holste... Westdeutsch... 0  
## 5 7.09 Dithmarschen Schleswig-Holste... Westdeutsch... 0  
## 6 7.36 Herzogtum Lauenbu... Schleswig-Holste... Westdeutsch... 0  
## 7 5.86 Nordfriesland Schleswig-Holste... Westdeutsch... 0  
## 8 5.90 Ostholstein Schleswig-Holste... Westdeutsch... 0  
## 9 6.16 Pinneberg  Schleswig-Holste... Westdeutsch... 0  
## 10 6.90 Plön     Schleswig-Holste... Westdeutsch... 0  
## # ... with 391 more rows
```

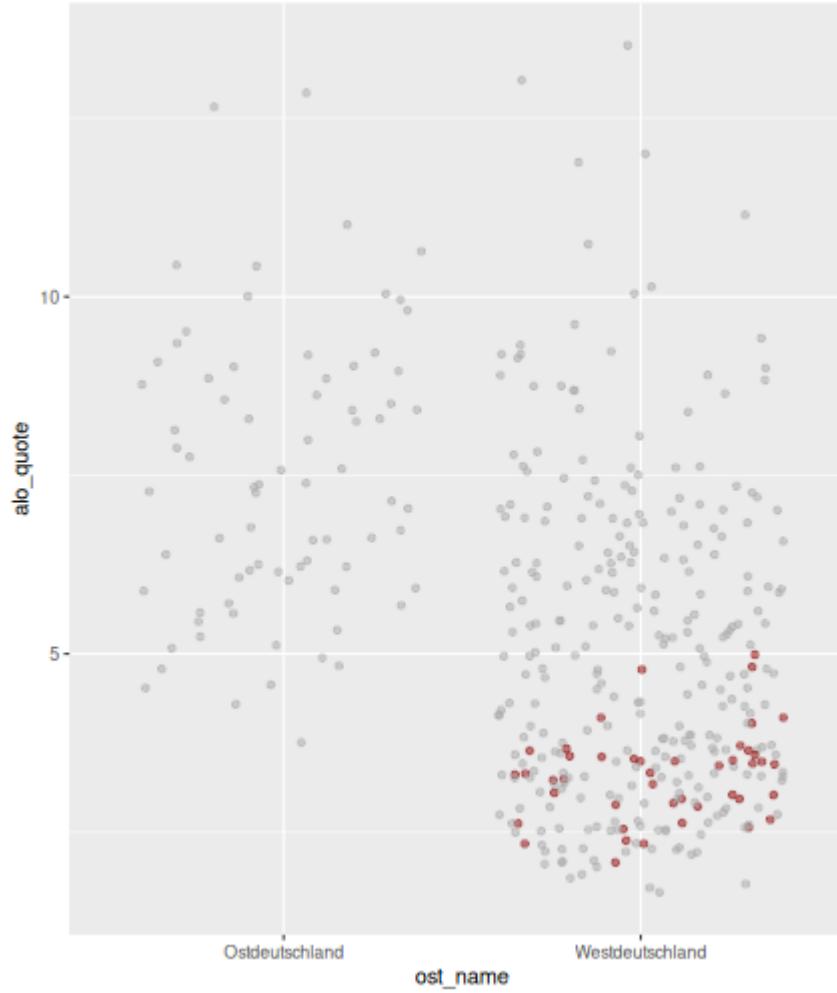
```
gesamtdaten %>%  
  select(alo_quote, landkreis_name, bundesland_name,  
         mutate(baden_wuerttemberg = as.factor(ifelse(bunde  
ggplot(aes(x = ost_name, y=alo_quote))
```



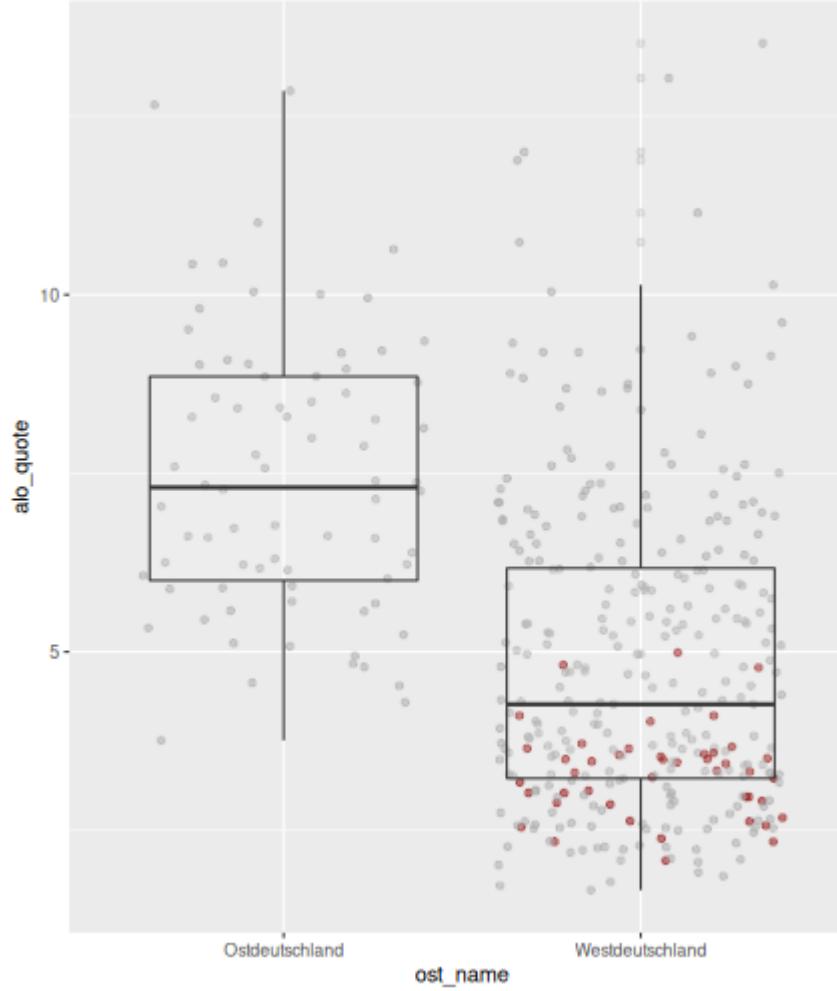
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
```



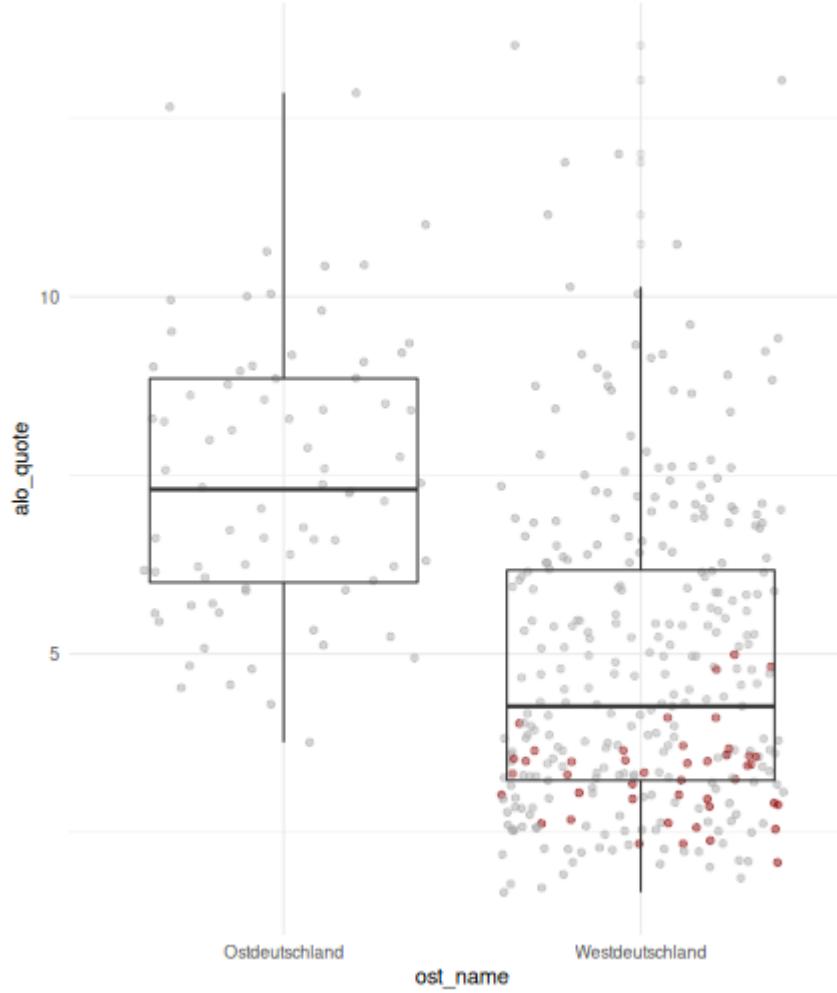
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity()
```



```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity() +
  geom_boxplot(alpha = 0.1)
```



```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity() +
  geom_boxplot(alpha = 0.1) +
  theme_minimal()
```



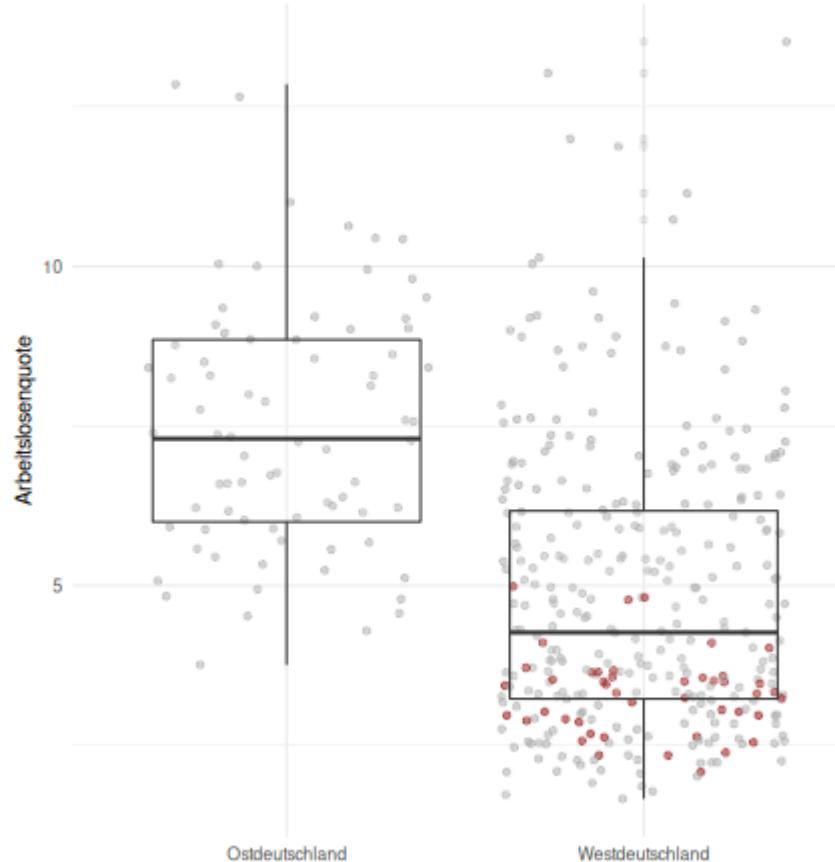
```

gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
  ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity() +
  geom_boxplot(alpha = 0.1) +
  theme_minimal() +
  labs(title = "Arbeitslosenquote in Deutschland",
       subtitle = "Eine Beobachtung repräsentiert ei
  x = "",
  y = "Arbeitslosenquote",
  caption = "Quelle: Daten der Agentur für Arbe

```

Arbeitslosenquote in Deutschland

Eine Beobachtung repräsentiert einen Landkreis, Baden-Württemberg rot eingefärbt



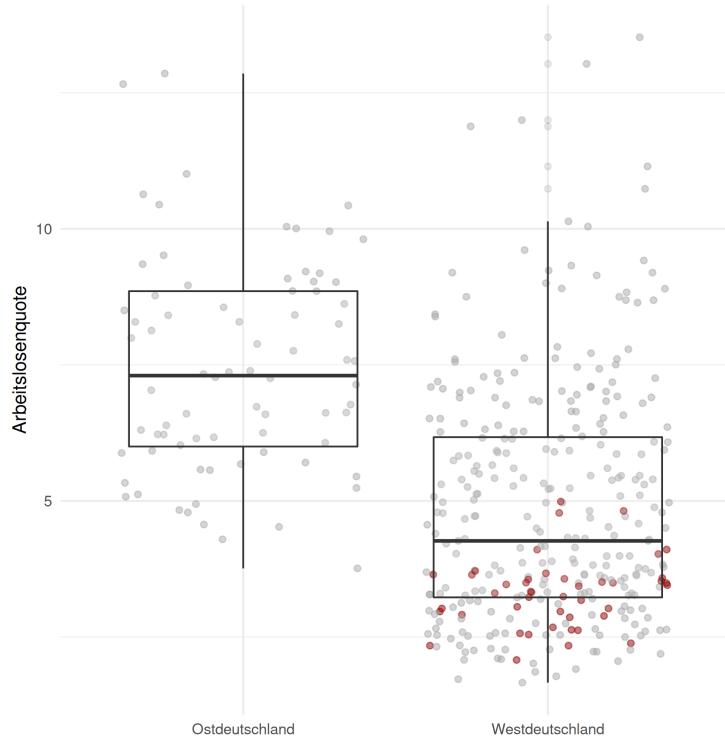
Quelle: Daten der Agentur für Arbeit aus dem Jahr 2017

Arbeitslosenquote

Beschreiben Sie das gezeigte Schaubild

Arbeitslosenquote in Deutschland

Eine Beobachtung repräsentiert einen Landkreis, Baden-Württemberg rot eingefärbt



Quelle: Daten der Agentur für Arbeit aus dem Jahr 2017

Arbeitslosenquote

Beschreibung des Schaubilds:

- ✚ Rote Datenpunkte Bade-Württemberg, fast alle unter dem Median in Westdeutschland
- ✚ Median in Westdeutschland deutlich geringer als in Ostdeutschland
- ✚ 75% Quantil in Westdeutschland entspricht (fast) 25% Quantil in Ostdeutschland
- ✚ Alle Landkreise unter 15% Arbeitslosenquote; Verglichen mit den europäischen Daten sehr gut

Bruttoinlandsprodukt pro Kopf

Es gibt deutliche regionale Unterschiede zwischen den Landkreisen. Doch ist dies auch beim BIP pro Kopf der Fall? Und war das schon immer so?

Wir betrachten das BIP pro Kopf über die Zeit für ost- und westdeutsche Landkreise!

Hier können wir sehen:

- + ob es auch regionale Unterschiede im BIP pro Kopf gibt
- + ob die regionalen Unterschiede schon längere Zeit bestehen
- + ob die regionalen Unterschiede sich vergrößern oder verkleinern

Bruttoinlandsprodukt pro Kopf

Das Bruttoinlandsprodukt stellt die wichtigste gesamtwirtschaftliche Kenngröße dar. Falls das BIP in einem Landkreis hoch ist könnte dies unter anderem daran liegen, dass

- ✚ viele Personen in diesem Landkreis erwerbstätig sind,
- ✚ oder das die Erwerbstätigen in Branchen mit hoher Produktivität arbeiten.

Falls der erste Punkt zutrifft sollte ein hohes BIP pro Kopf (berechnet als BIP pro **Einwohner**) tendenziell auch mit einer niedrigeren Arbeitslosenquote einhergehen.

Bruttoinlandsprodukt pro Kopf

Das Bruttoinlandsprodukt stellt die wichtigste gesamtwirtschaftliche Kenngröße dar. Falls das BIP in einem Landkreis hoch ist könnte dies unter anderem daran liegen, dass

- ✚ viele Personen in diesem Landkreis erwerbstätig sind,
- ✚ oder das die Erwerbstätigen in Branchen mit hoher Produktivität arbeiten.

Falls der erste Punkt zutrifft sollte ein hohes BIP pro Kopf (berechnet als BIP pro **Einwohner**) tendenziell auch mit einer niedrigeren Arbeitslosenquote einhergehen.

Beschreiben und interpretieren Sie das gezeigte Schaubild.

```
options(scipen = 5)
```

```
options(scipen = 5)
bip_zeitreihe_namen
```

```
## # A tibble: 9,975 x 8
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>    <dbl>      <dbl> <fct>
## 1 08111            1992 3.20e10  593628    53930. Baden-Württemb...
## 2 08115            1992 1.20e10  343190    34900. Baden-Württemb...
## 3 08116            1992 1.22e10  487370    25057. Baden-Württemb...
## 4 08117            1992 5.12e 9  248688    20586. Baden-Württemb...
## 5 08118            1992 1.15e10  475248    24170. Baden-Württemb...
## 6 08119            1992 8.49e 9  389670    21782. Baden-Württemb...
## 7 08121            1992 4.21e 9  118566    35543. Baden-Württemb...
## 8 08125            1992 6.11e 9  283163    21585. Baden-Württemb...
## 9 08126            1992 2.29e 9  96072     23797. Baden-Württemb...
## 10 08127            1992 3.49e 9  169617    20551. Baden-Württemb...
## # ... with 9,965 more rows, and 2 more variables: landkreis_name <fct>,
## #       ost_name <chr>
```

```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 )
```

```
## # A tibble: 7,182 x 8
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>    <dbl>      <dbl> <fct>
## 1 08111            2000 3.43e10  571528     59965. Baden-Württemb...
## 2 08115            2000 1.38e10  359476     38259. Baden-Württemb...
## 3 08116            2000 1.44e10  492914     29224. Baden-Württemb...
## 4 08117            2000 6.12e 9  253970     24094. Baden-Württemb...
## 5 08118            2000 1.44e10  492014     29350. Baden-Württemb...
## 6 08119            2000 1.04e10  403830     25799. Baden-Württemb...
## 7 08121            2000 5.31e 9  115590     45970. Baden-Württemb...
## 8 08125            2000 8.57e 9  316406     27075. Baden-Württemb...
## 9 08126            2000 3.12e 9  106494     29305. Baden-Württemb...
## 10 08127            2000 4.60e 9  184222     24984. Baden-Württemb...
## # ... with 7,172 more rows, and 2 more variables: landkreis_name <fct>,
## #       ost_name <chr>
```

```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr)
```

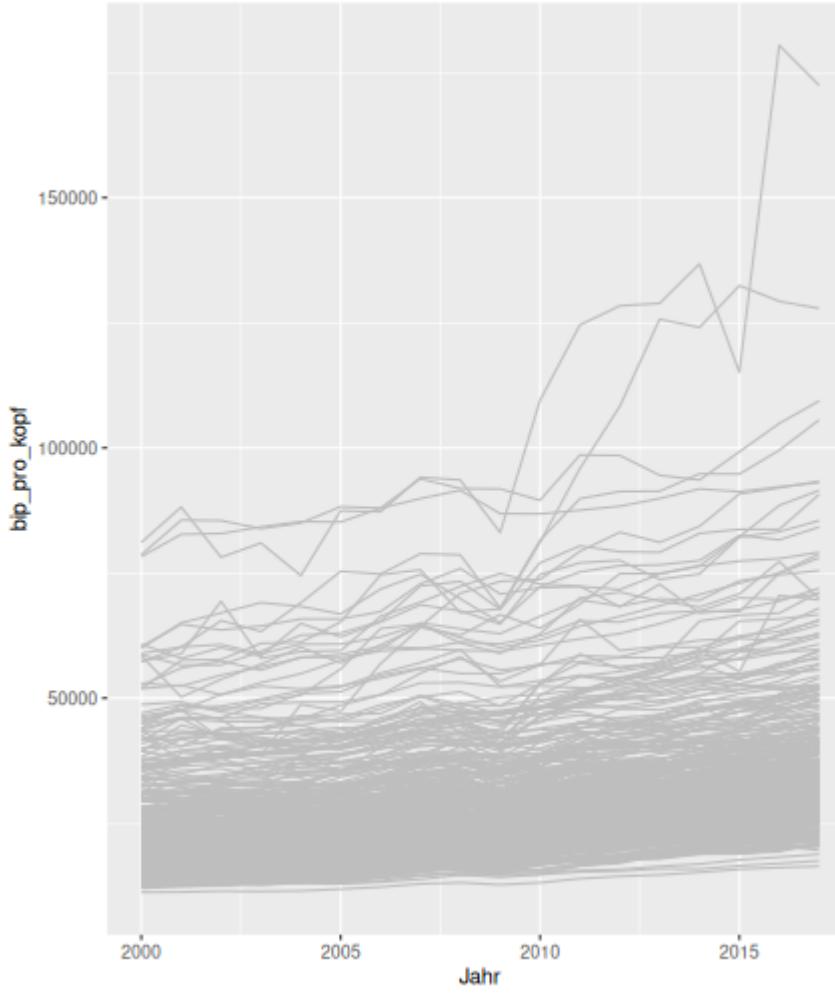
```
## # A tibble: 7,182 x 8
## # Groups:   ost_name, Jahr [36]
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>      <dbl>        <dbl> <fct>
## 1 08111          2000 3.43e10  571528       59965. Baden-Württemb...
## 2 08115          2000 1.38e10  359476       38259. Baden-Württemb...
## 3 08116          2000 1.44e10  492914       29224. Baden-Württemb...
## 4 08117          2000 6.12e 9   253970       24094. Baden-Württemb...
## 5 08118          2000 1.44e10  492014       29350. Baden-Württemb...
## 6 08119          2000 1.04e10  403830       25799. Baden-Württemb...
## 7 08121          2000 5.31e 9   115590       45970. Baden-Württemb...
## 8 08125          2000 8.57e 9   316406       27075. Baden-Württemb...
## 9 08126          2000 3.12e 9   106494       29305. Baden-Württemb...
## 10 08127         2000 4.60e 9   184222       24984. Baden-Württemb...
## # ... with 7,172 more rows, and 2 more variables: landkreis_name <fct>,
## #   ost_name <chr>
```

```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_
```

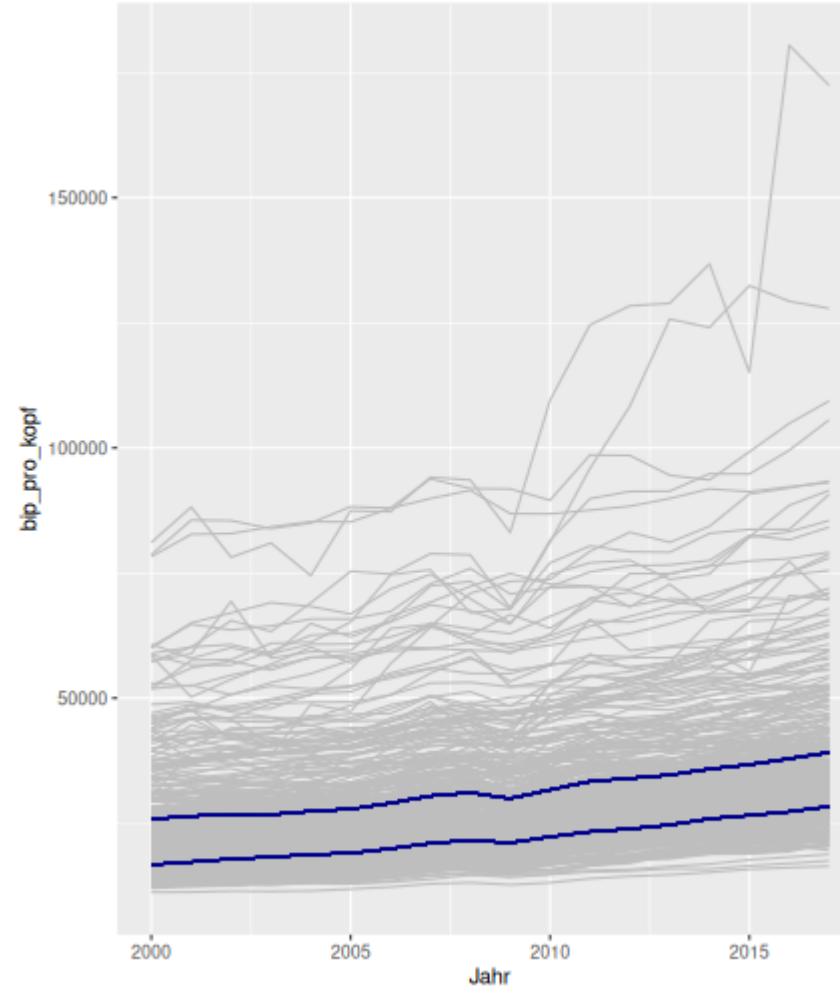
```
## # A tibble: 7,182 x 10
## # Groups:   ost_name, Jahr [36]
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>      <dbl>        <dbl> <fct>
## 1 08111           2000 3.43e10  571528       59965. Baden-Württemb...
## 2 08115           2000 1.38e10  359476       38259. Baden-Württemb...
## 3 08116           2000 1.44e10  492914       29224. Baden-Württemb...
## 4 08117           2000 6.12e 9   253970       24094. Baden-Württemb...
## 5 08118           2000 1.44e10  492014       29350. Baden-Württemb...
## 6 08119           2000 1.04e10  403830       25799. Baden-Württemb...
## 7 08121           2000 5.31e 9   115590       45970. Baden-Württemb...
## 8 08125           2000 8.57e 9   316406       27075. Baden-Württemb...
## 9 08126           2000 3.12e 9   106494       29305. Baden-Württemb...
## 10 08127          2000 4.60e 9   184222       24984. Baden-Württemb...
## # ... with 7,172 more rows, and 4 more variables: landkreis_name <fct>,
## #   ost_name <chr>, durchschnitt <dbl>, ulm <dbl>
```

```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_
ggplot()
```

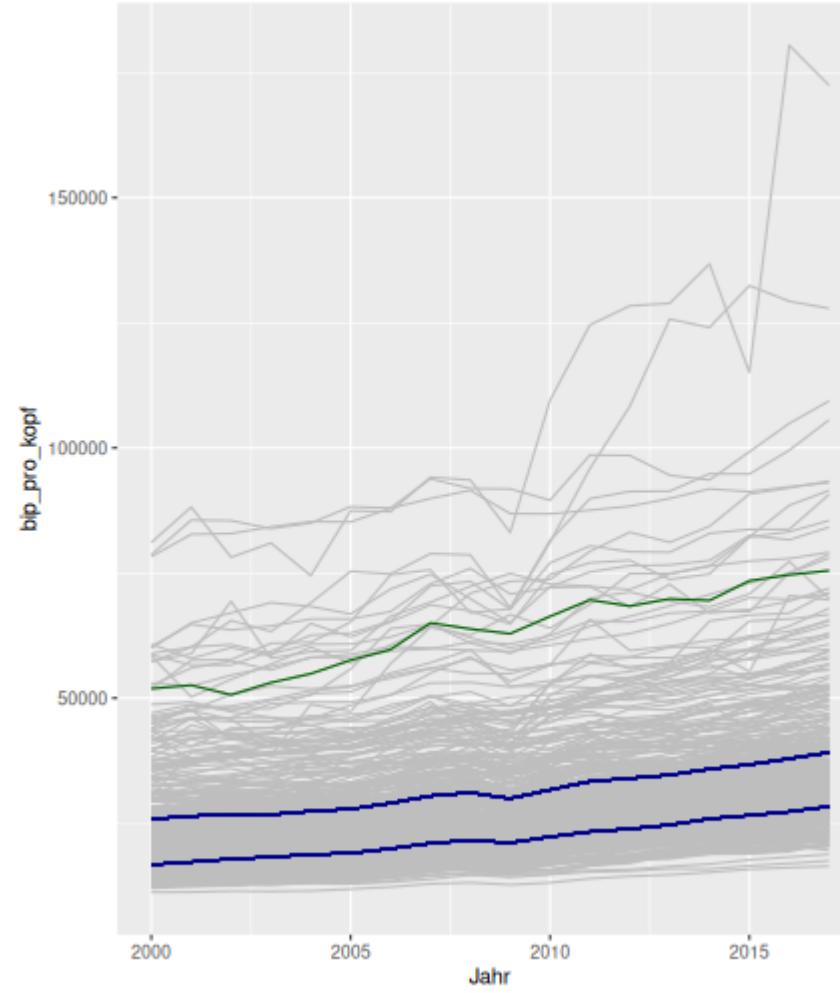
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_-
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
```



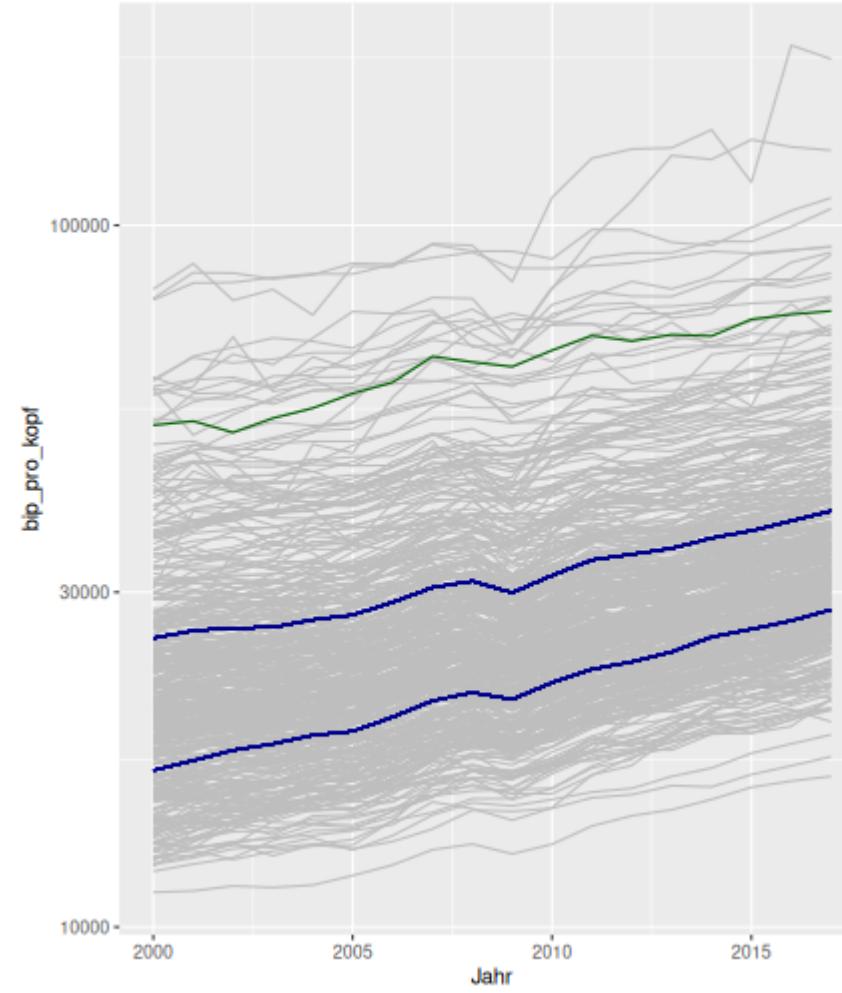
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0),
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
```



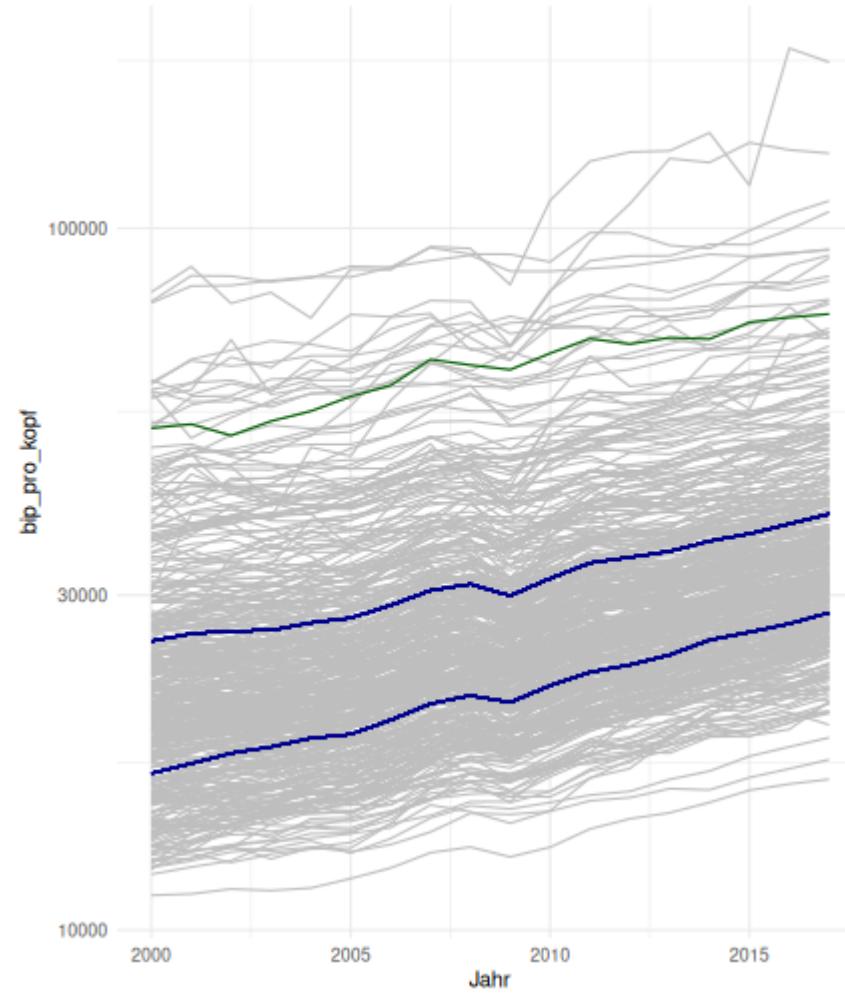
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_
```



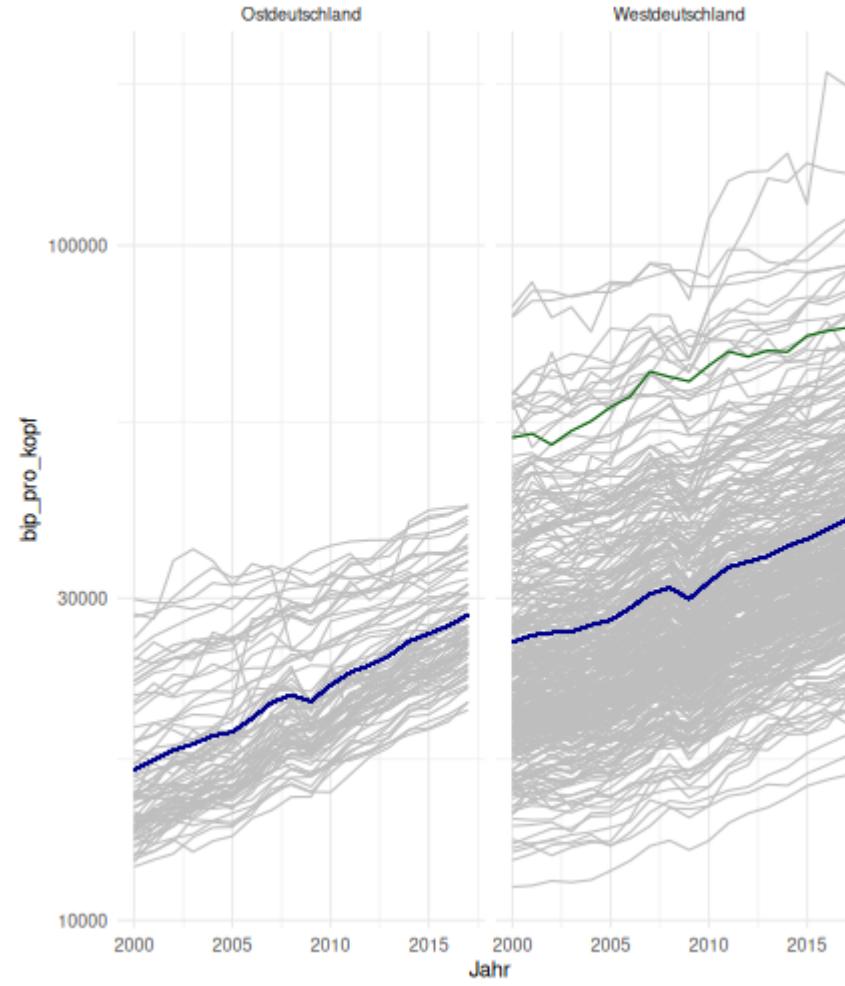
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  geom_line(aes(x = Jahr, y = ulm, group = Regionals
  scale_y_continuous(trans = "log10")
```



```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  geom_line(aes(x = Jahr, y = ulm, group = Regionals
  scale_y_continuous(trans = "log10") +
  theme_minimal()
```



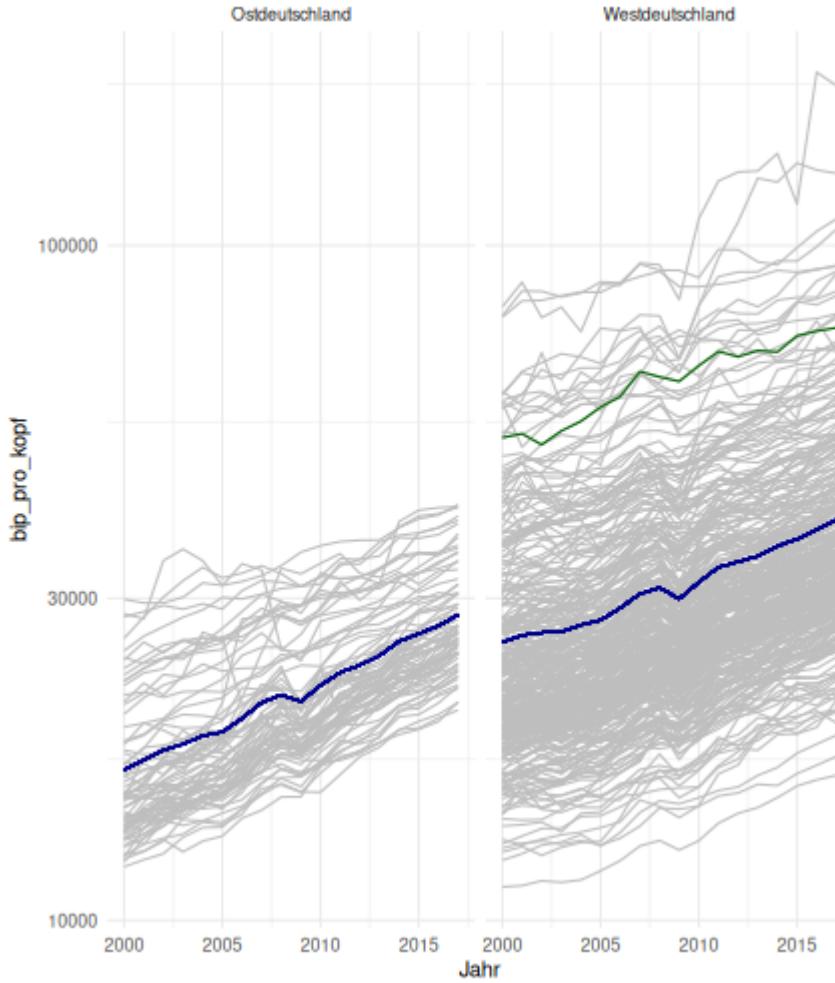
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  geom_line(aes(x = Jahr, y = ulm, group = Regionals
  scale_y_continuous(trans = "log10") +
  theme_minimal() +
  facet_wrap(ost_name ~ .)
```



```

options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  geom_line(aes(x = Jahr, y = ulm, group = Regionals
  scale_y_continuous(trans = "log10") +
  theme_minimal() +
  facet_wrap(ost_name ~ .) +
  theme(legend.position = "none")

```

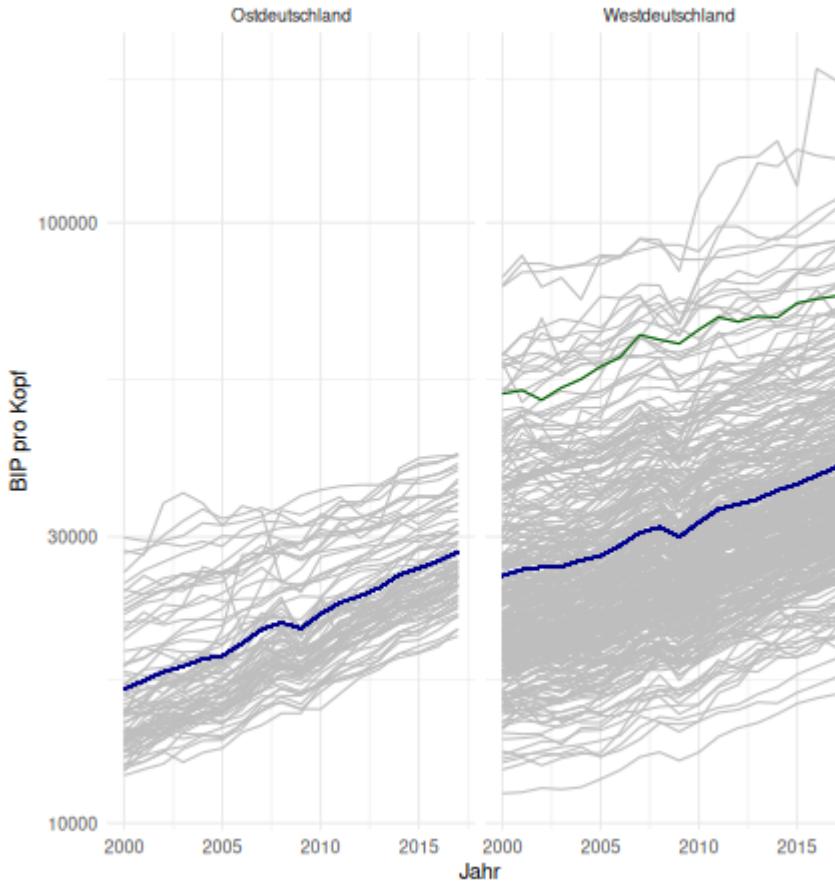


```

options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, NA))
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  geom_line(aes(x = Jahr, y = ulm, group = Regionals
  scale_y_continuous(trans = "log10") +
  theme_minimal() +
  facet_wrap(ost_name ~ .) +
  theme(legend.position = "none") +
  labs(title = "Ein Vergleich des BIP pro Kopf von ost- und westdeutschen Land",
       subtitle = "Durchschnittswerte in Dunkelblau, Ulm in Dunkelgrün",
       x = "Jahr",
       y = "BIP pro Kopf")

```

Ein Vergleich des BIP pro Kopf von ost- und westdeutschen Land
Durchschnittswerte in Dunkelblau, Ulm in Dunkelgrün



Quelle: Daten der Statistischen Ämter der Länder und des Bundes.

Bruttoinlandsprodukt pro Kopf

Beschreibung:

- + Logarithmische Skalierung der y-Achse
- + Das Niveau des BIP pro Kopf ist in den ostdeutschen Bundesländern deutlich niedriger als in den westdeutschen.
- + Stadtkreis Ulm hat ein sehr hohes BIP pro Kopf, auch im Zeitablauf
- + Das BIP Pro Kopf nimmt im Zeitablauf in Ostdeutschland zu, doch erreicht es mit durchschnittlich 28338€ den Wert, welchen die westdeutschen Landkreise durchschnittlich in 2006 hatten!
- + In 2008/2009 gibt es überall einen Einbruch beim BIP pro Kopf, jedoch scheint dieser in den ostdeutschen Bundesländern nicht so stark gewesen zu sein

Bruttoinlandsprodukt pro Kopf

Beschreibung:

- + Logarithmische Skalierung der y-Achse
- + Das Niveau des BIP pro Kopf ist in den ostdeutschen Bundesländern deutlich niedriger als in den westdeutschen.
- + Stadtkreis Ulm hat ein sehr hohes BIP pro Kopf, auch im Zeitablauf
- + Das BIP Pro Kopf nimmt im Zeitablauf in Ostdeutschland zu, doch erreicht es mit durchschnittlich 28338€ den Wert, welchen die westdeutschen Landkreise durchschnittlich in 2006 hatten!
- + In 2008/2009 gibt es überall einen Einbruch beim BIP pro Kopf, jedoch scheint dieser in den ostdeutschen Bundesländern nicht so stark gewesen zu sein

Interpretation:

- + Eine Wachstumsprozess im BIP pro Kopf findet in allen Landkreisen statt, jedoch gibt es für die ostdeutschen Landkreise, welche deutlich niedriger gestartet sind, keinen erkennbaren Anpassungsprozess in Form eines schnelleren Wachstums
- + Wir sehen auch keinen Anpassungsprozess der Landkreise in Westdeutschland
- + Fraglich ist, ob wir hier mit einem Anpassungsprozess von strukturschwachen Landkreisen überhaupt rechnen sollten

Wachstum des BIP

Paneldaten beim BIP, d.h. wir können:

- ✚ Das Wachstum des BIP
- ✚ Für alle Landkreise in Deutschland
- ✚ Seit 2000 bis 2017

visualisieren.

Wachstum des BIP

Paneldaten beim BIP, d.h. wir können:

- + Das Wachstum des BIP
- + Für alle Landkreise in Deutschland
- + Seit 2000 bis 2017

visualisieren.

| Gab es einen Anpassungsprozess des BIP pro Kopf?

```
bip_zeitreihe_namen
```

```
## # A tibble: 9,975 x 8
##   Regionalschlues... Jahr   bip_einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>     <dbl>       <dbl>      <dbl> <fct>
## 1 08111            1992 3.20e10    593628    53930. Baden-Württemb...
## 2 08115            1992 1.20e10    343190    34900. Baden-Württemb...
## 3 08116            1992 1.22e10    487370    25057. Baden-Württemb...
## 4 08117            1992 5.12e 9    248688    20586. Baden-Württemb...
## 5 08118            1992 1.15e10    475248    24170. Baden-Württemb...
## 6 08119            1992 8.49e 9    389670    21782. Baden-Württemb...
## 7 08121            1992 4.21e 9    118566    35543. Baden-Württemb...
## 8 08125            1992 6.11e 9    283163    21585. Baden-Württemb...
## 9 08126            1992 2.29e 9    96072     23797. Baden-Württemb...
## 10 08127            1992 3.49e 9   169617    20551. Baden-Württemb...
## # ... with 9,965 more rows, and 2 more variables: landkreis_name <fct>,
## #   ost_name <chr>
```

```
bip_zeitreihe_namen %>%  
  group_by(Regionalschluessel)
```

```
## # A tibble: 9,975 x 8  
## # Groups: Regionalschlüssel [399]  
##   Regionalschlüssel Jahr   bip_einwohner bip_pro_kopf bundesland_name  
##   <chr>           <dbl>     <dbl>       <dbl>      <dbl> <fct>  
## 1 08111            1992 3.20e10    593628    53930. Baden-Württemb...  
## 2 08115            1992 1.20e10    343190    34900. Baden-Württemb...  
## 3 08116            1992 1.22e10    487370    25057. Baden-Württemb...  
## 4 08117            1992 5.12e 9    248688    20586. Baden-Württemb...  
## 5 08118            1992 1.15e10    475248    24170. Baden-Württemb...  
## 6 08119            1992 8.49e 9    389670    21782. Baden-Württemb...  
## 7 08121            1992 4.21e 9    118566    35543. Baden-Württemb...  
## 8 08125            1992 6.11e 9    283163    21585. Baden-Württemb...  
## 9 08126            1992 2.29e 9    96072     23797. Baden-Württemb...  
## 10 08127           1992 3.49e 9    169617    20551. Baden-Württemb...  
## # ... with 9,965 more rows, and 2 more variables: landkreis_name <fct>,  
## #   ost_name <chr>
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr)

## # A tibble: 9,975 x 8
## # Groups:   Regionalschluessel [399]
##   Regionalschlues... Jahr   bip einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl> <dbl>      <dbl>       <dbl> <fct>
## 1 01001        1992 2.64e9    86642     30468. Schleswig-Hols...
## 2 01001        1994 2.73e9    86287     31622. Schleswig-Hols...
## 3 01001        1995 2.74e9    85506     32033. Schleswig-Hols...
## 4 01001        1996 2.69e9    84499     31835. Schleswig-Hols...
## 5 01001        1997 2.85e9    83344     34208. Schleswig-Hols...
## 6 01001        1998 3.00e9    82112     36488. Schleswig-Hols...
## 7 01001        1999 2.83e9    81276     34779. Schleswig-Hols...
## 8 01001        2000 2.62e9    80758     32428. Schleswig-Hols...
## 9 01001        2001 2.64e9    80489     32854. Schleswig-Hols...
## 10 01001       2002 2.71e9    80414     33734. Schleswig-Hols...
## # ... with 9,965 more rows, and 2 more variables: landkreis_name <fct>,
## #       ost_name <chr>
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)

## # A tibble: 9,975 x 9
## # Groups:   Regionalschluessel [399]
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl>  <dbl>      <dbl>      <dbl> <fct>
## 1 01001        1992 2.64e9    86642     30468. Schleswig-Hols...
## 2 01001        1994 2.73e9    86287     31622. Schleswig-Hols...
## 3 01001        1995 2.74e9    85506     32033. Schleswig-Hols...
## 4 01001        1996 2.69e9    84499     31835. Schleswig-Hols...
## 5 01001        1997 2.85e9    83344     34208. Schleswig-Hols...
## 6 01001        1998 3.00e9    82112     36488. Schleswig-Hols...
## 7 01001        1999 2.83e9    81276     34779. Schleswig-Hols...
## 8 01001        2000 2.62e9    80758     32428. Schleswig-Hols...
## 9 01001        2001 2.64e9    80489     32854. Schleswig-Hols...
## 10 01001       2002 2.71e9    80414     33734. Schleswig-Hols...
## # ... with 9,965 more rows, and 3 more variables: landkreis_name <fct>,
## #       ost_name <chr>, bip_wachstum <dbl>
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup()

## # A tibble: 9,975 x 9
##   Regionalschlues... Jahr   bip einwohner bip_pro_kopf bundesland_name
##   <chr>          <dbl> <dbl>    <dbl>      <dbl> <fct>
## 1 01001           1992 2.64e9    86642     30468. Schleswig-Hols...
## 2 01001           1994 2.73e9    86287     31622. Schleswig-Hols...
## 3 01001           1995 2.74e9    85506     32033. Schleswig-Hols...
## 4 01001           1996 2.69e9    84499     31835. Schleswig-Hols...
## 5 01001           1997 2.85e9    83344     34208. Schleswig-Hols...
## 6 01001           1998 3.00e9    82112     36488. Schleswig-Hols...
## 7 01001           1999 2.83e9    81276     34779. Schleswig-Hols...
## 8 01001           2000 2.62e9    80758     32428. Schleswig-Hols...
## 9 01001           2001 2.64e9    80489     32854. Schleswig-Hols...
## 10 01001          2002 2.71e9    80414     33734. Schleswig-Hols...
## # ... with 9,965 more rows, and 3 more variables: landkreis_name <fct>,
## #       ost_name <chr>, bip_wachstum <dbl>
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup() %>%
  group_by(ost_name, Jahr)
```

```
## # A tibble: 9,975 x 9
## # Groups:   ost_name, Jahr [50]
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl>    <dbl>      <dbl>       <dbl> <fct>
## 1 01001          1992 2.64e9    86642     30468. Schleswig-Hols...
## 2 01001          1994 2.73e9    86287     31622. Schleswig-Hols...
## 3 01001          1995 2.74e9    85506     32033. Schleswig-Hols...
## 4 01001          1996 2.69e9    84499     31835. Schleswig-Hols...
## 5 01001          1997 2.85e9    83344     34208. Schleswig-Hols...
## 6 01001          1998 3.00e9    82112     36488. Schleswig-Hols...
## 7 01001          1999 2.83e9    81276     34779. Schleswig-Hols...
## 8 01001          2000 2.62e9    80758     32428. Schleswig-Hols...
## 9 01001          2001 2.64e9    80489     32854. Schleswig-Hols...
## 10 01001         2002 2.71e9    80414     33734. Schleswig-Hols...
## # ... with 9,965 more rows, and 3 more variables: landkreis_name <fct>,
## #   ost_name <chr>, bip_wachstum <dbl>
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE)
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum
```

```
## # A tibble: 9,975 x 10
## # Groups:   ost_name, Jahr [50]
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl>    <dbl>      <dbl>      <dbl> <fct>
## 1 01001          1992 2.64e9    86642     30468. Schleswig-Hols...
## 2 01001          1994 2.73e9    86287     31622. Schleswig-Hols...
## 3 01001          1995 2.74e9    85506     32033. Schleswig-Hols...
## 4 01001          1996 2.69e9    84499     31835. Schleswig-Hols...
## 5 01001          1997 2.85e9    83344     34208. Schleswig-Hols...
## 6 01001          1998 3.00e9    82112     36488. Schleswig-Hols...
## 7 01001          1999 2.83e9    81276     34779. Schleswig-Hols...
## 8 01001          2000 2.62e9    80758     32428. Schleswig-Hols...
## 9 01001          2001 2.64e9    80489     32854. Schleswig-Hols...
## 10 01001         2002 2.71e9    80414     33734. Schleswig-Hols...
## # ... with 9,965 more rows, and 4 more variables: landkreis_name <fct>,
## #   ost_name <chr>, bip_wachstum <dbl>, durchschnitt <dbl>
```

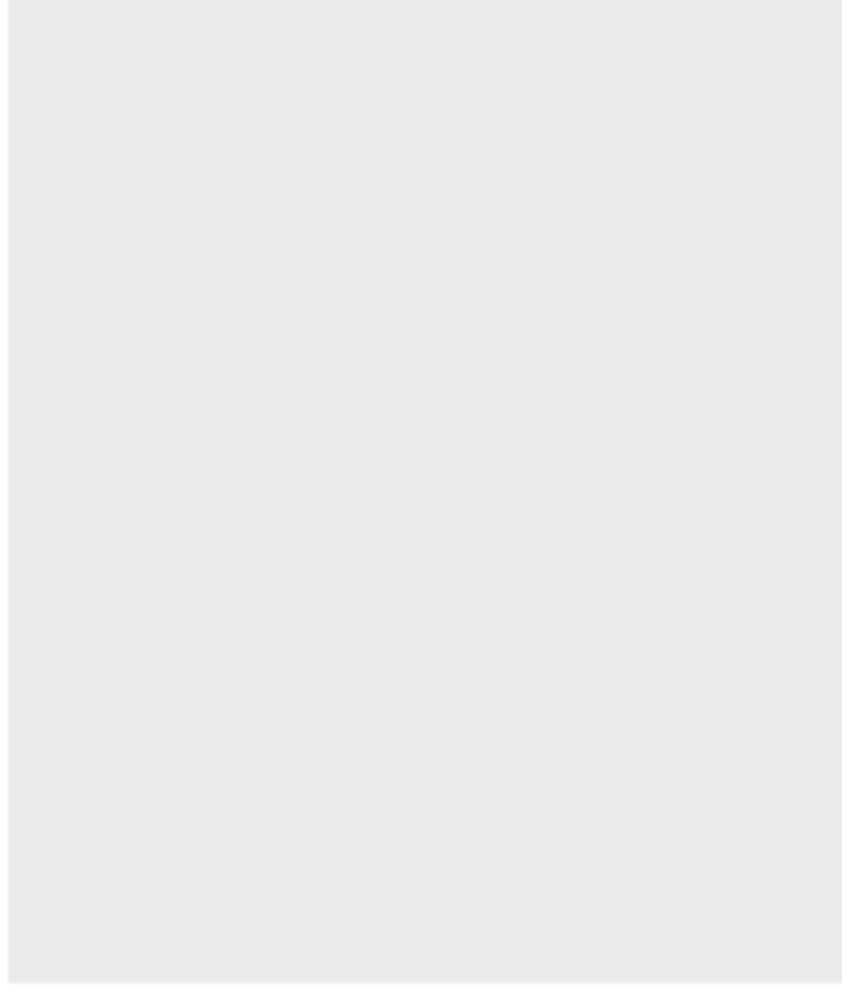
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 )
```

```
## # A tibble: 7,182 x 10
## # Groups:   ost_name, Jahr [36]
##   Regionalschlues... Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl>    <dbl>      <dbl>       <dbl> <fct>
## 1 01001        2000 2.62e9    80758     32428. Schleswig-Hols...
## 2 01001        2001 2.64e9    80489     32854. Schleswig-Hols...
## 3 01001        2002 2.71e9    80414     33734. Schleswig-Hols...
## 4 01001        2003 2.74e9    80538     33976. Schleswig-Hols...
## 5 01001        2004 2.93e9    80783     36231. Schleswig-Hols...
## 6 01001        2005 3.01e9    80892     37258. Schleswig-Hols...
## 7 01001        2006 3.13e9    81052     38676. Schleswig-Hols...
## 8 01001        2007 3.18e9    81634     38931. Schleswig-Hols...
## 9 01001        2008 3.25e9    82403     39456. Schleswig-Hols...
## 10 01001       2009 3.13e9    82478     37920. Schleswig-Hols...
## # ... with 7,172 more rows, and 4 more variables: landkreis_name <fct>,
## #   ost_name <chr>, bip_wachstum <dbl>, durchschnitt <dbl>
```

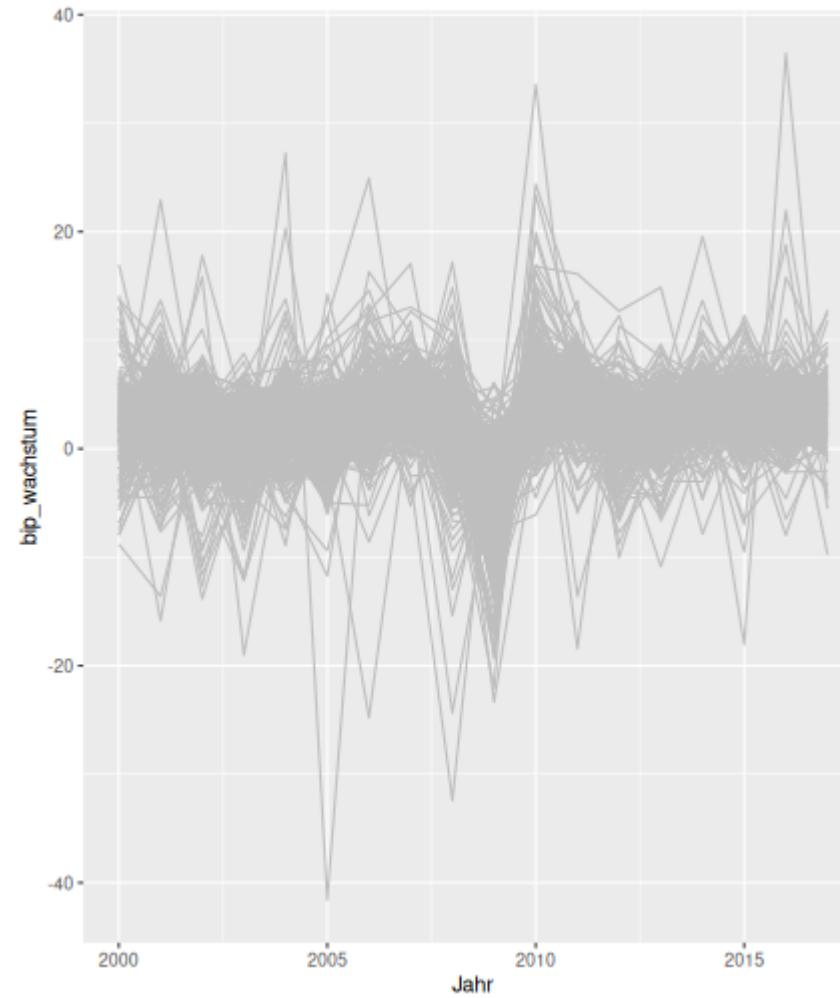
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot()
```



```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
```

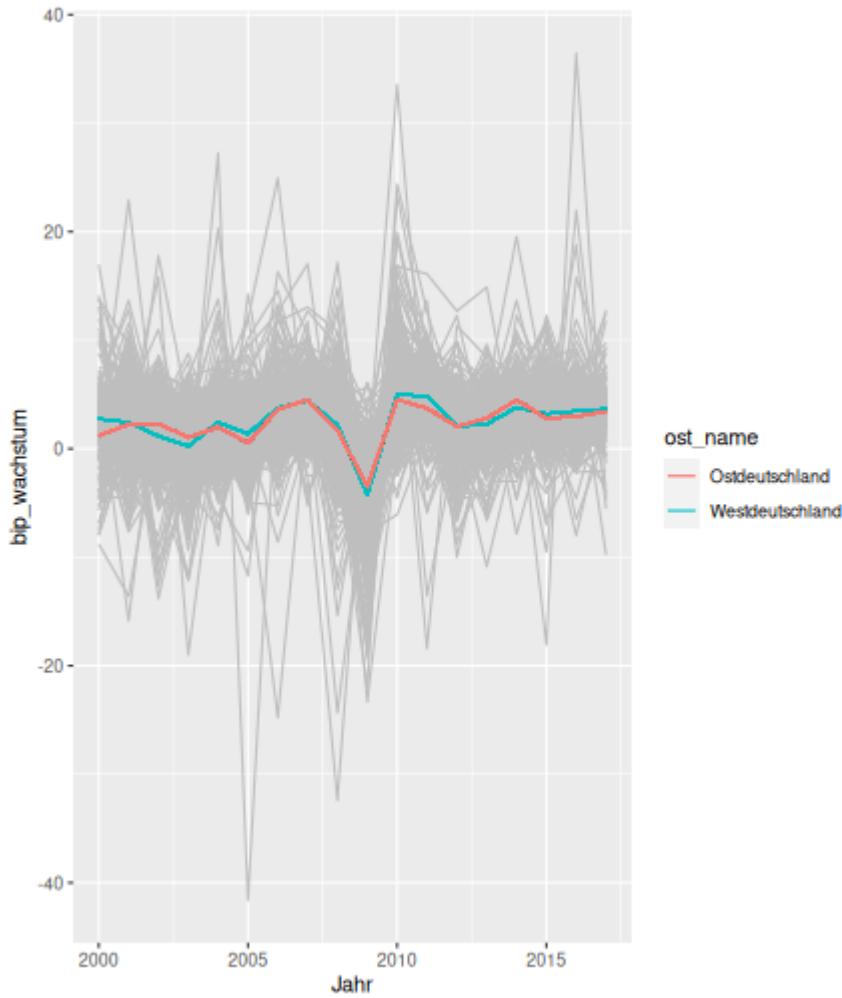


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =

```

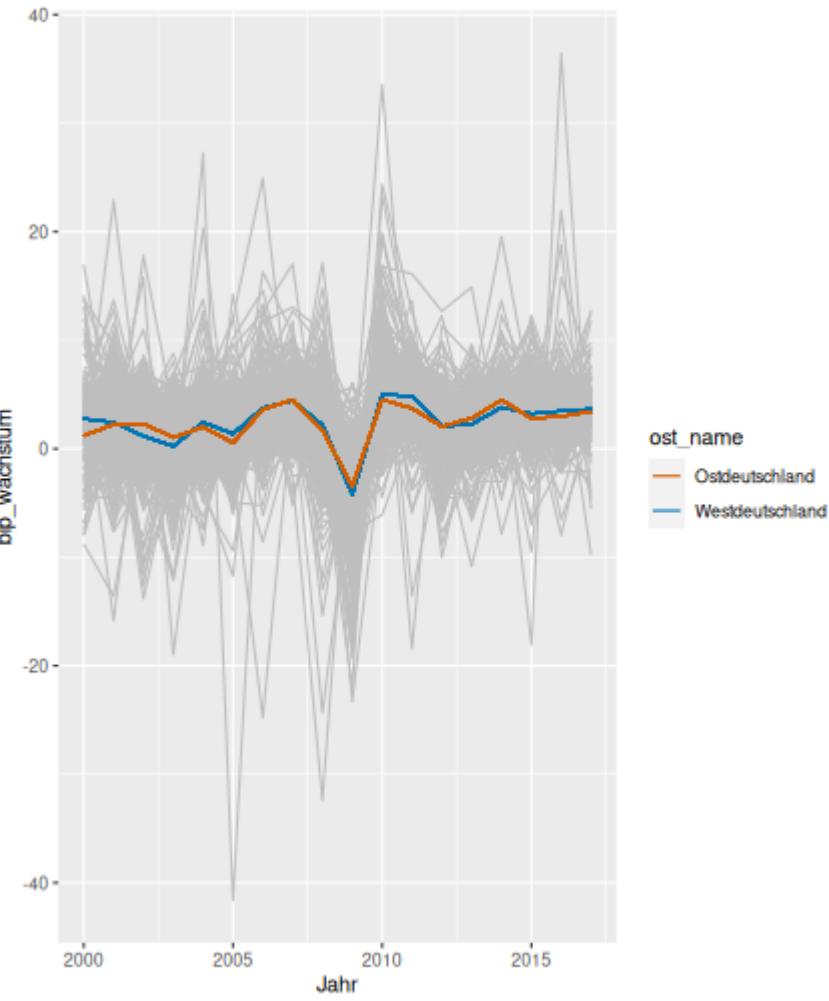


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"))

```

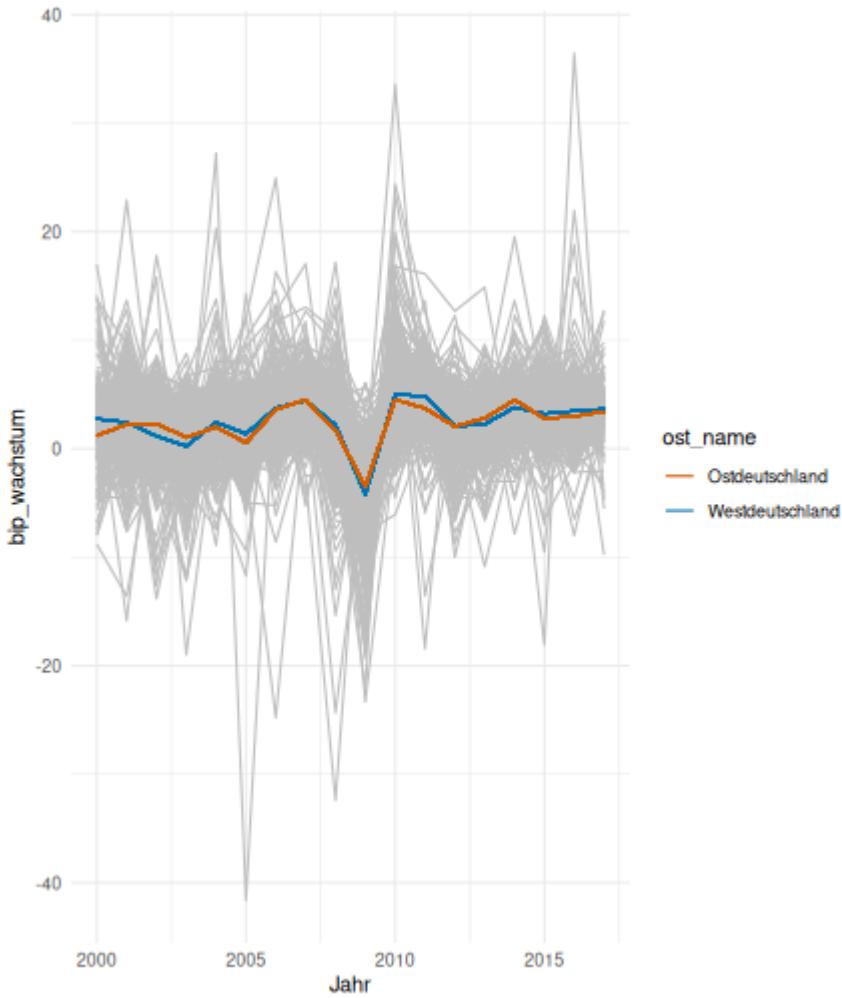


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"))
  theme_minimal()

```

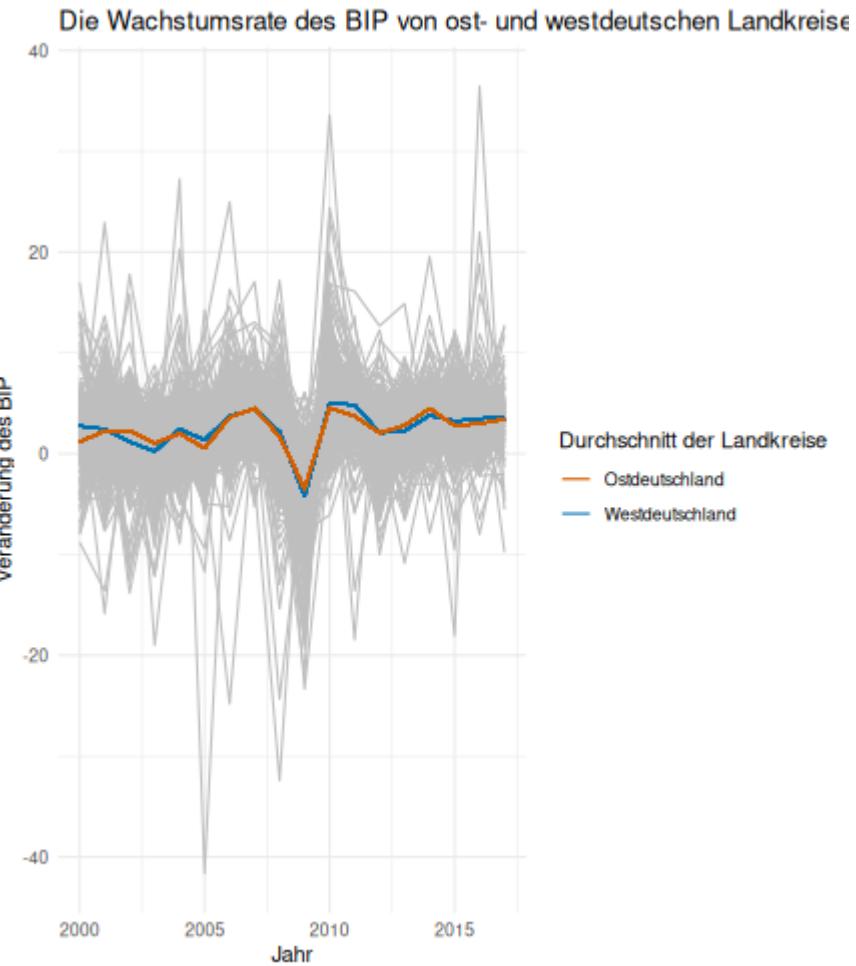


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"))
  theme_minimal() +
  labs(color = "Durchschnitt der Landkreise",
       title = "Die Wachstumsrate des BIP von ost- und westdeutschen Landkreisen",
       caption = "Quelle: Daten der Statistischen Ämter der Länder und des Bundes.",
       x = "Jahr",
       y = "Veränderung des BIP")

```

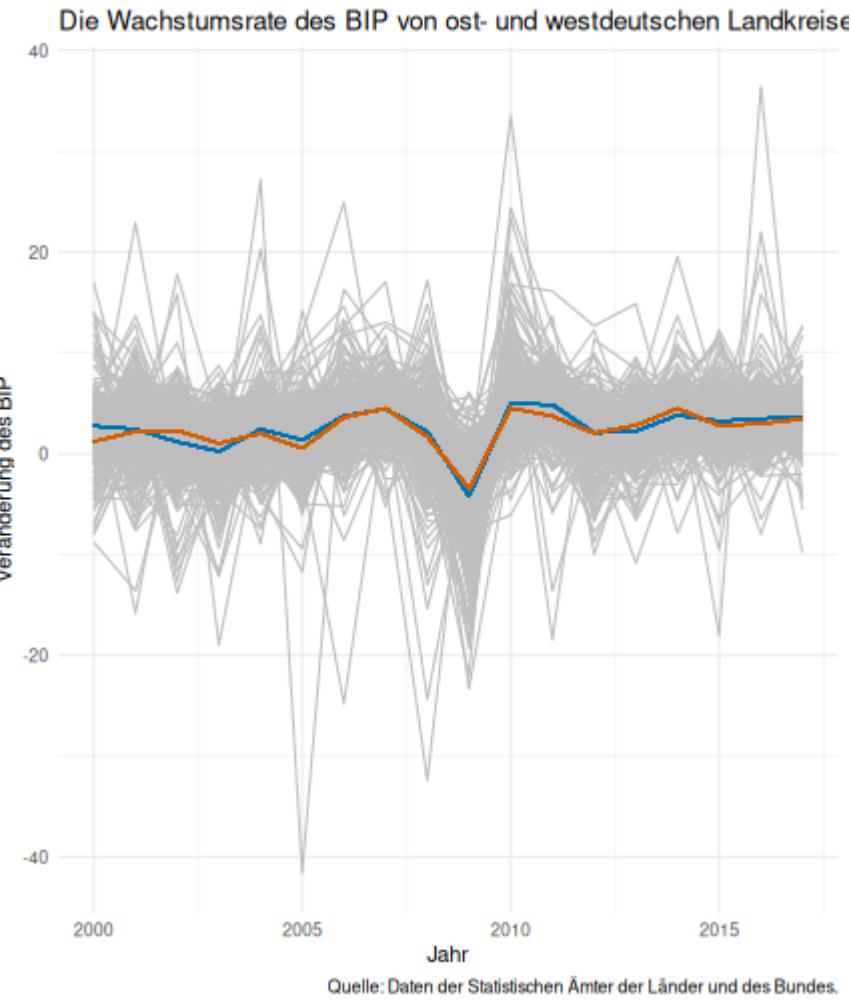


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"))
  theme_minimal() +
  labs(color = "Durchschnitt der Landkreise",
       title = "Die Wachstumsrate des BIP von ost- u
  caption = "Quelle: Daten der Statistischen Äm
  x = "Jahr",
  y = "Veränderung des BIP") +
  theme(legend.position = "none")

```



```

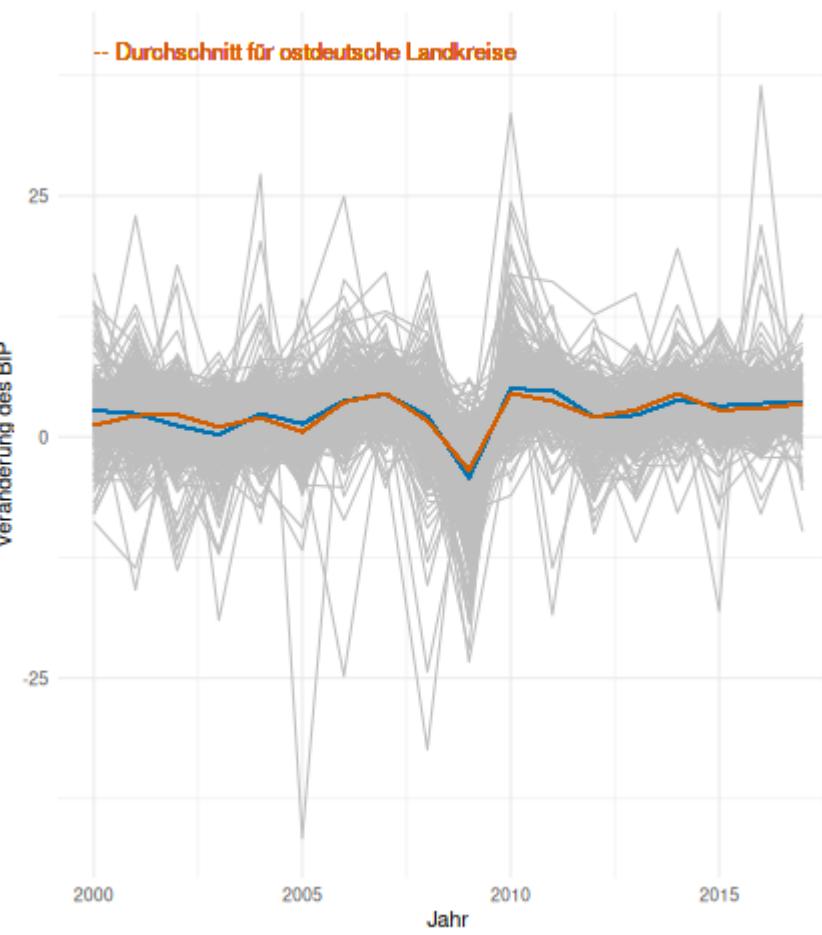
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"))
  theme_minimal() +
  labs(color = "Durchschnitt der Landkreise",
       title = "Die Wachstumsrate des BIP von ost- u
  caption = "Quelle: Daten der Statistischen Äm
  x = "Jahr",
  y = "Veränderung des BIP") +
  theme(legend.position = "none") +
  geom_text(aes(x=2000, y=40, label = "-- Durchschni

```

Die Wachstumsrate des BIP von ost- und westdeutschen Landkreisen

-- Durchschnitt für ostdeutsche Landkreise



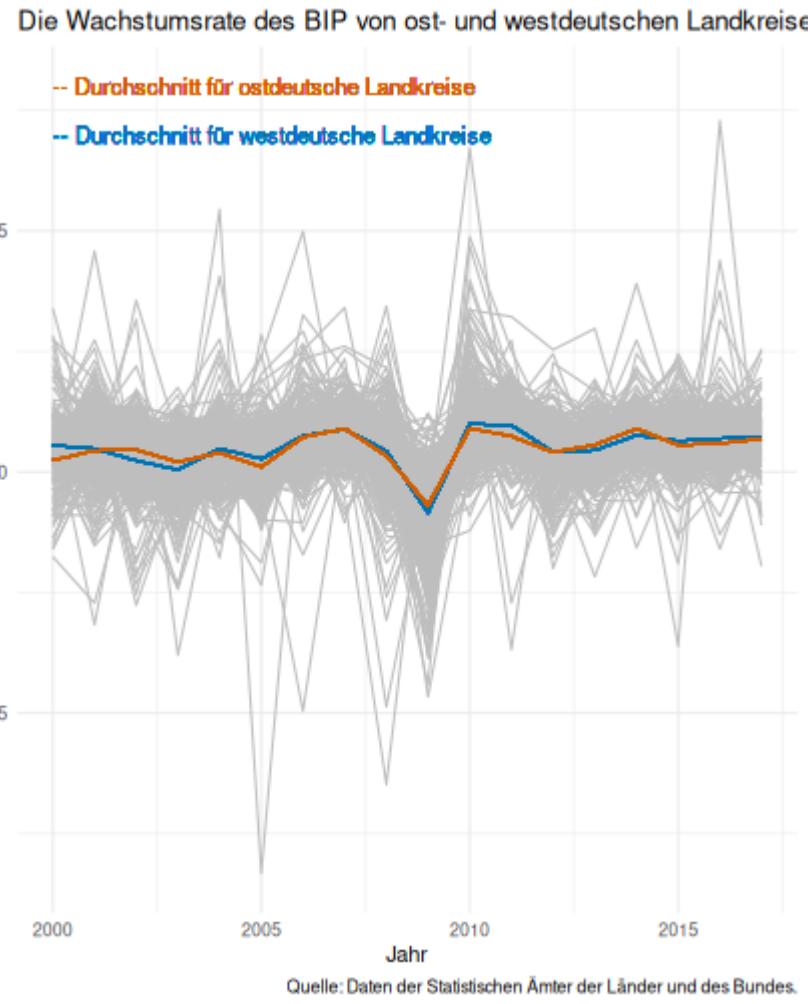
Quelle: Daten der Statistischen Ämter der Länder und des Bundes.

```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_wachstum = 100*(bip - lag(bip)) / bip)
ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_wachstum, na.rm=TRUE))

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_wachstum, group =
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"))
  theme_minimal() +
  labs(color = "Durchschnitt der Landkreise",
       title = "Die Wachstumsrate des BIP von ost- u
  caption = "Quelle: Daten der Statistischen Äm
  x = "Jahr",
  y = "Veränderung des BIP") +
  theme(legend.position = "none") +
  geom_text(aes(x=2000, y=40, label = "-- Durchschni
  geom_text(aes(x=2000, y=35, label = "-- Durchschni

```



Wachstum des BIP pro Kopf

Beschreibung:

- + Im Durchschnitt sehr ähnliche Wachstumsraten
- + Immer wieder vereinzelt sehr hohe Wachstumsraten pro Landkreis
 - + Hängt vermutlich mit großen Projekten auf Landkreisebene zusammen
- + Der Einbruch in der Finanzkrise ist sowohl bei ost- als auch westdeutschen Landkreisen zu sehen

Interpretation:

- + Es findet keine Anpassung des BIP pro Kopf über die Zeit statt
- + Die Gelder durch den Soli-Ausgleich führen nicht zu der (erhofften) starken Aufholjagd
- + Ostdeutsche Landkreise haben sich stark entwickelt
 - + Diese Entwicklung sollte jedoch nicht absolut, sondern relativ zu westdeutschen Landkreisen betrachtet werden

Wachstum des BIP pro Kopf

Beschreibung:

- + Im Durchschnitt sehr ähnliche Wachstumsraten
- + Immer wieder vereinzelt sehr hohe Wachstumsraten pro Landkreis
 - + Hängt vermutlich mit großen Projekten auf Landkreisebene zusammen
- + Der Einbruch in der Finanzkrise ist sowohl bei ost- als auch westdeutschen Landkreisen zu sehen

Interpretation:

- + Es findet keine Anpassung des BIP pro Kopf über die Zeit statt
- + Die Gelder durch den Soli-Ausgleich führen nicht zu der (erhofften) starken Aufholjagd
- + Ostdeutsche Landkreise haben sich stark entwickelt
 - + Diese Entwicklung sollte jedoch nicht absolut, sondern relativ zu westdeutschen Landkreisen betrachtet werden

Es ist kein Anpassungsprozess ersichtlich, dafür sind die Wachstumsraten zu ähnlich.

Verteilung des BIP pro Kopf in 2017

Bisherige Grafiken:

- ✚ Punktewolke + Boxplot zeigt die Verteilung
- ✚ Liniendiagramm zeigt die Entwicklung

Verteilung des BIP pro Kopf in 2017

Bisherige Grafiken:

- + Punktewolke + Boxplot zeigt die Verteilung
- + Liniendiagramm zeigt die Entwicklung

Alternative Darstellungen der Verteilung:

- + Histogramm (nächste Folie)
- + Kerndichteschätzer (übernächste Folie)

Alternative Darstellung der Entwicklung:

- + Small multiples (siehe ausführliche Case-Study)
- + Slopechart (siehe z.B. [Data Vizualisation von Claus Wilke](#) mit [Code hier](#))

gesamtdaten

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  filter(bip_pro_kopf<100000)
```

```
## # A tibble: 395 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>          <dbl> <fct>           <chr>      <fct>
## 1 01001            4512 Flensburg     01        Schleswig-Hols...
## 2 01002            12345 Kiel          01        Schleswig-Hols...
## 3 01003            9692 Lübeck         01        Schleswig-Hols...
## 4 01004            3836 Neumünster    01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen   01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland  01        Schleswig-Hols...
## 8 01055            5748 Ostholstein    01        Schleswig-Hols...
## 9 01056            8599 Pinneberg     01        Schleswig-Hols...
## 10 01057            3264 Plön          01        Schleswig-Hols...
## # ... with 385 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name)

## # A tibble: 395 x 14
## # Groups:   ost_name [2]
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg   01        Schleswig-Hols...
## 10 01057           3264 Plön        01        Schleswig-Hols...
## # ... with 385 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

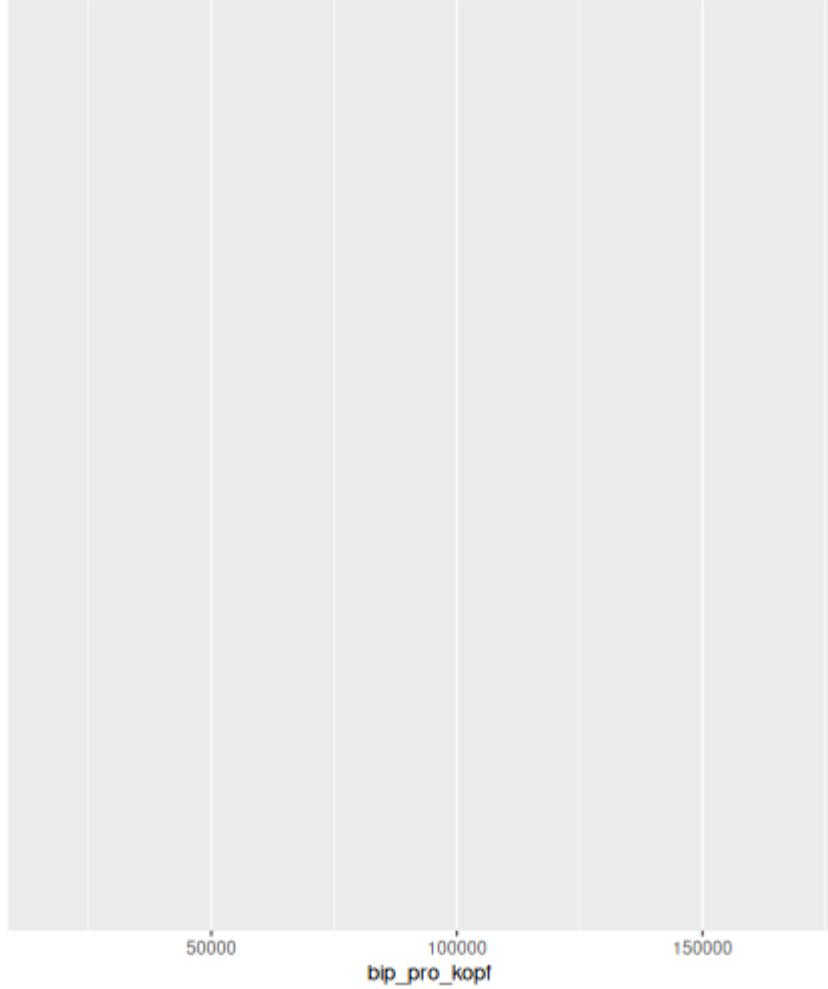
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf))
```

```
## # A tibble: 2 x 2
##   ost_name      durchschnitt
##   <chr>          <dbl>
## 1 Ostdeutschland    28338.
## 2 Westdeutschland   38021.
```

```
gesamtdataen %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel
```

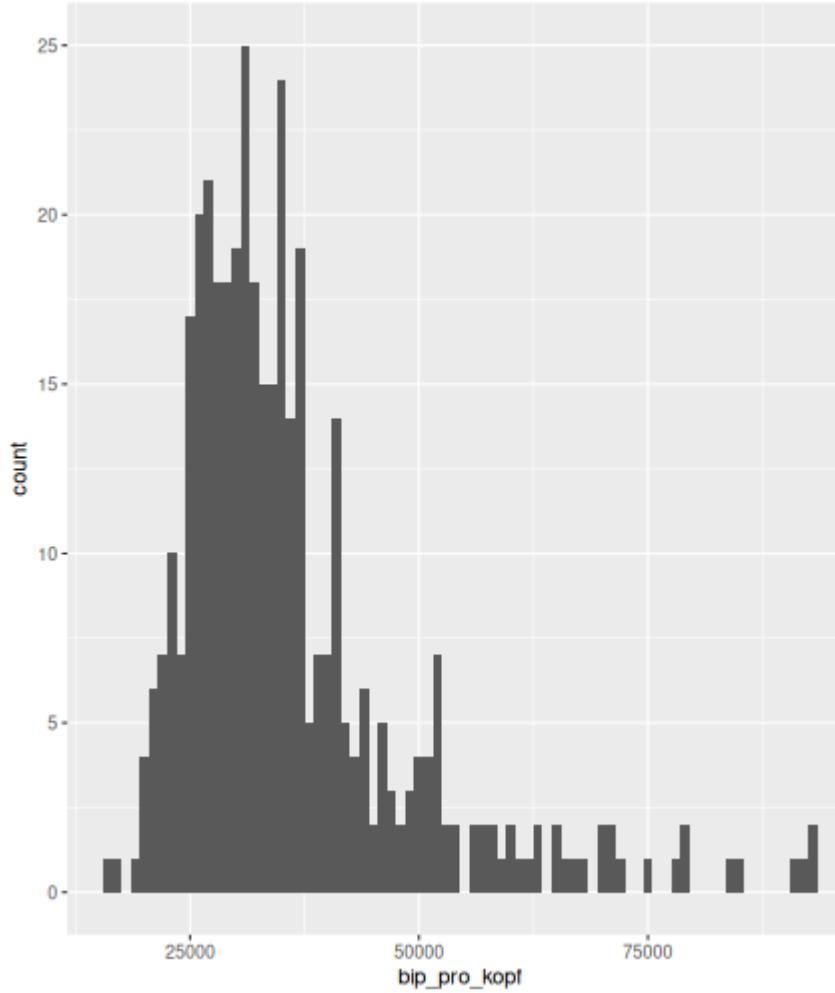
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf))
```



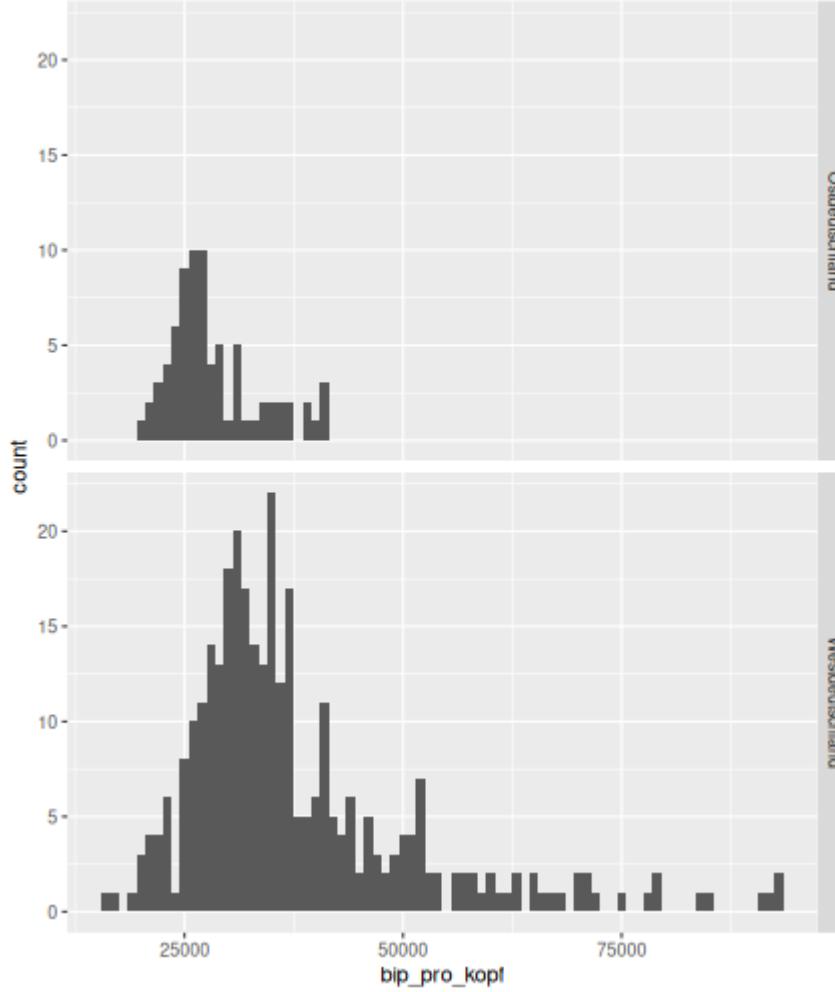
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_
```



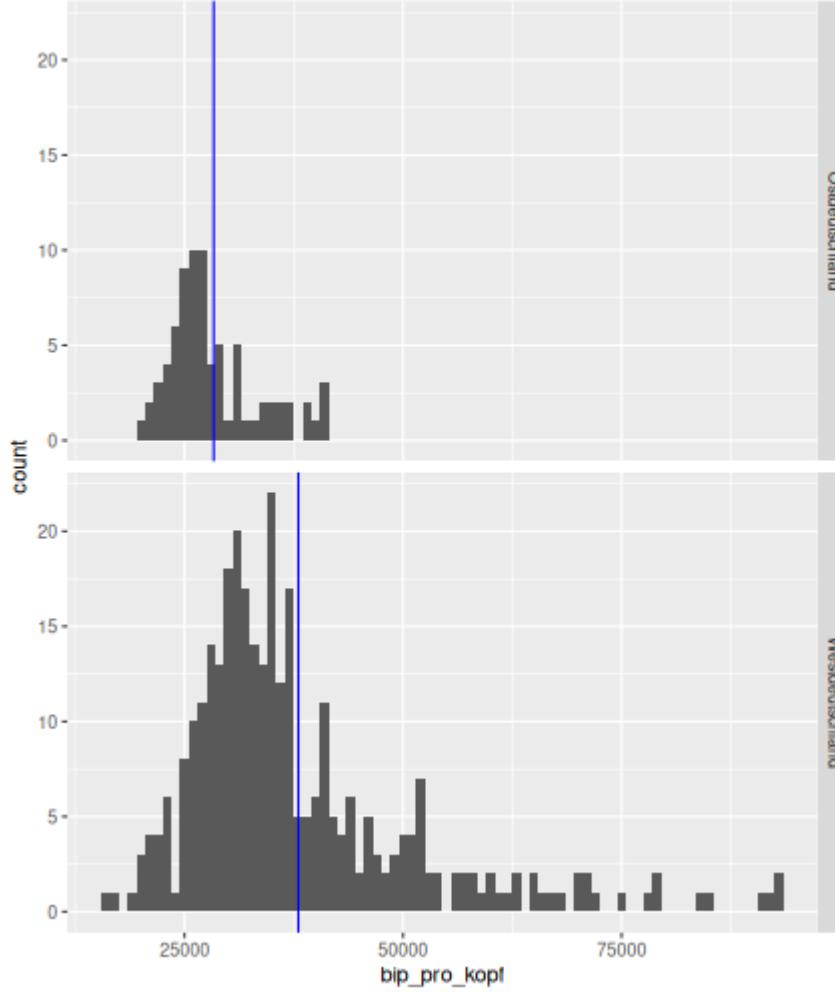
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_
```



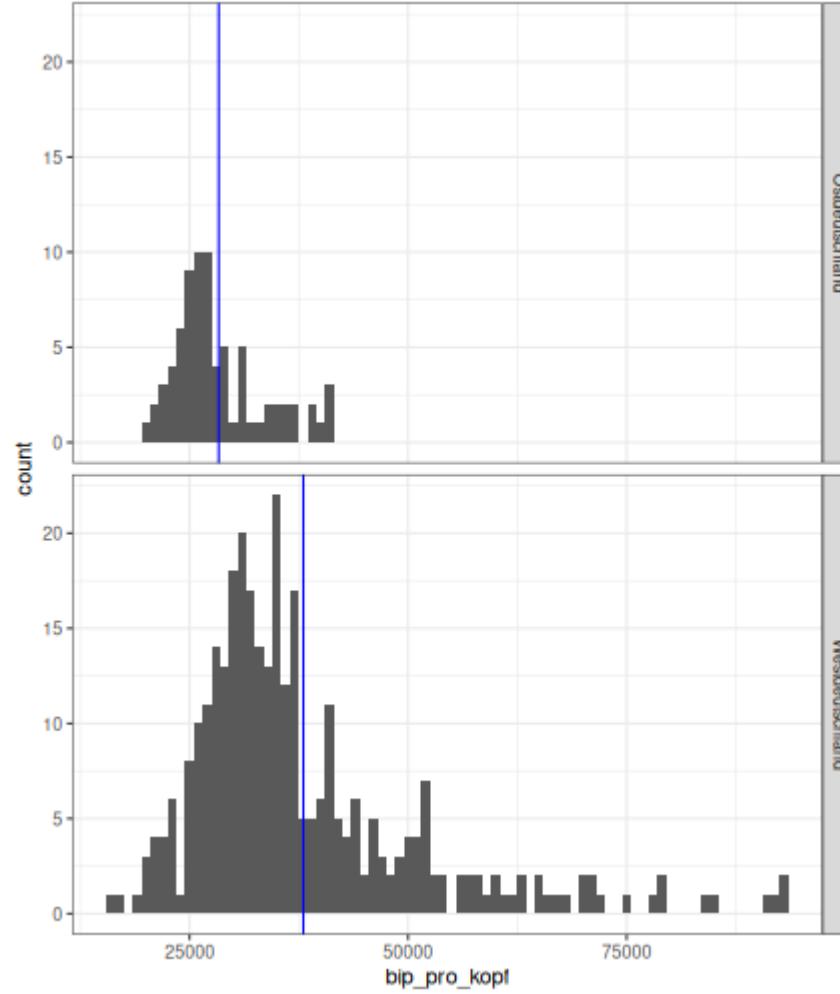
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf<100000), bins=50) +
  facet_grid(ost_name~.) +
  geom_vline(data = mittel, aes(xintercept = durchschnitt), color="blue")
```



```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf<100000), binwidth = 1000) +
  facet_grid(ost_name~.) +
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
  theme_bw()
```

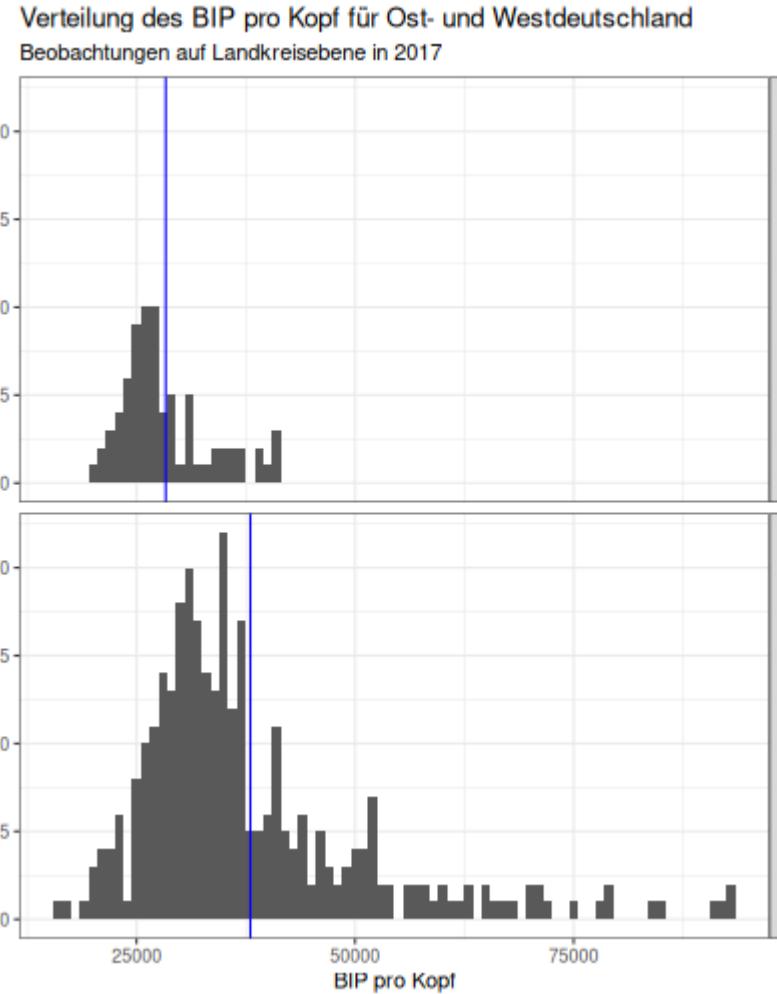


```

gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf)) %>%
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf<100000),
                 facet_grid(ost_name~.) +
  geom_vline(data = mittel, aes(xintercept = durchschnitt),
             theme_bw() +
  labs(title = "Verteilung des BIP pro Kopf für Ost- und Westdeutschland",
       subtitle = "Beobachtungen auf Landkreisebene",
       x = "BIP pro Kopf",
       y = "Anzahl an Beobachtungen")

```



gesamtdaten

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  filter(bip_pro_kopf<100000)
```

```
## # A tibble: 395 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>          <dbl> <fct>           <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg   01        Schleswig-Hols...
## 10 01057            3264 Plön        01        Schleswig-Hols...
## # ... with 385 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name)

## # A tibble: 395 x 14
## # Groups:   ost_name [2]
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg   01        Schleswig-Hols...
## 10 01057           3264 Plön        01        Schleswig-Hols...
## # ... with 385 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), median
```

```
## # A tibble: 2 x 3
##   ost_name      durchschnitt median
##   <chr>          <dbl>     <dbl>
## 1 Ostdeutschland    28338.  26775.
## 2 Westdeutschland   38021.  34579.
```

```
gesamtdataen %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2
```

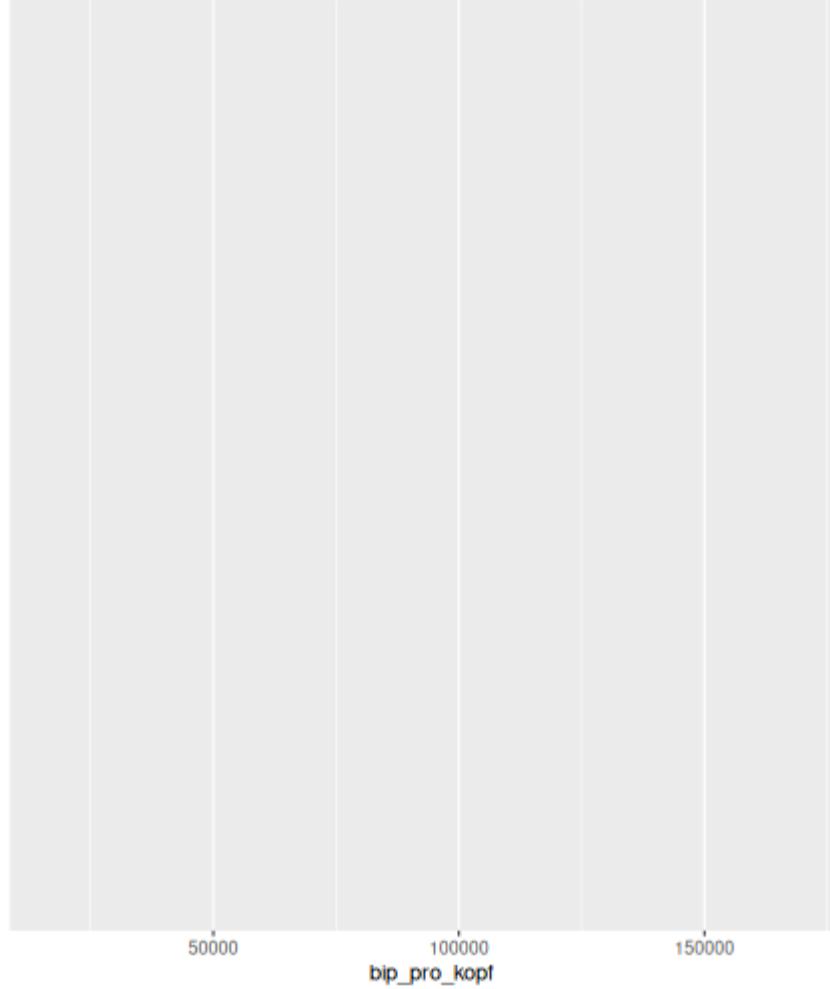
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten
```

```
## # A tibble: 401 x 14
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster  01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein  01        Schleswig-Hols...
## 9 01056            8599 Pinneberg   01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 9 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>
```

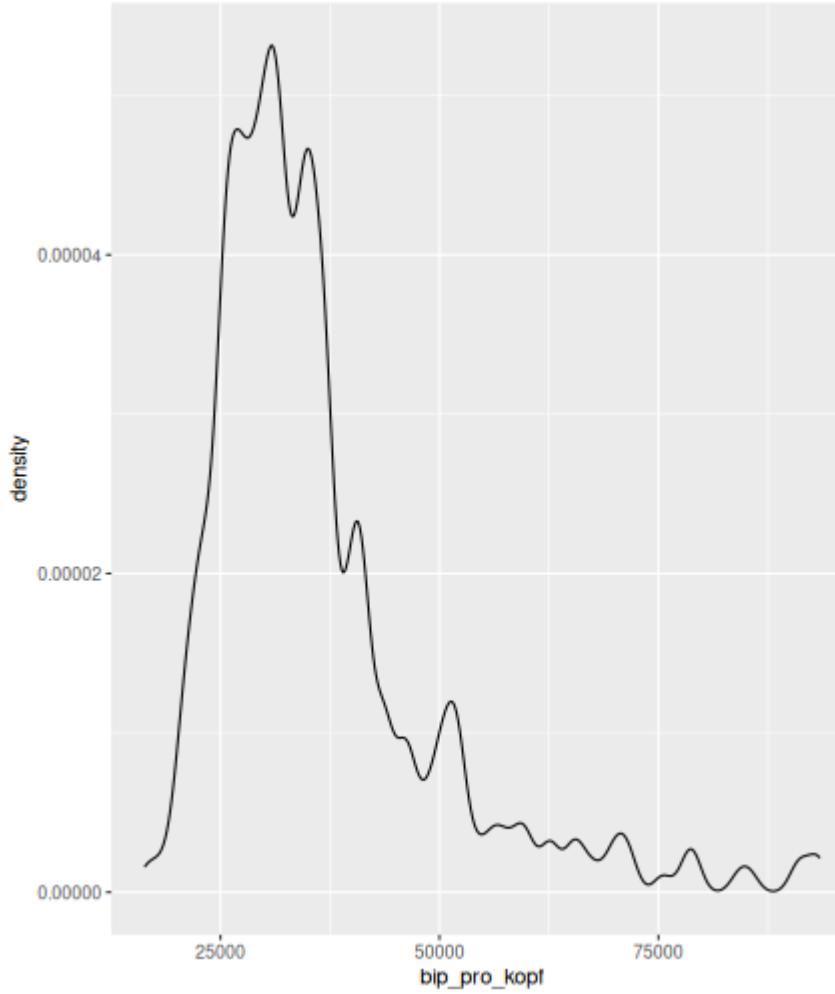
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf))
```



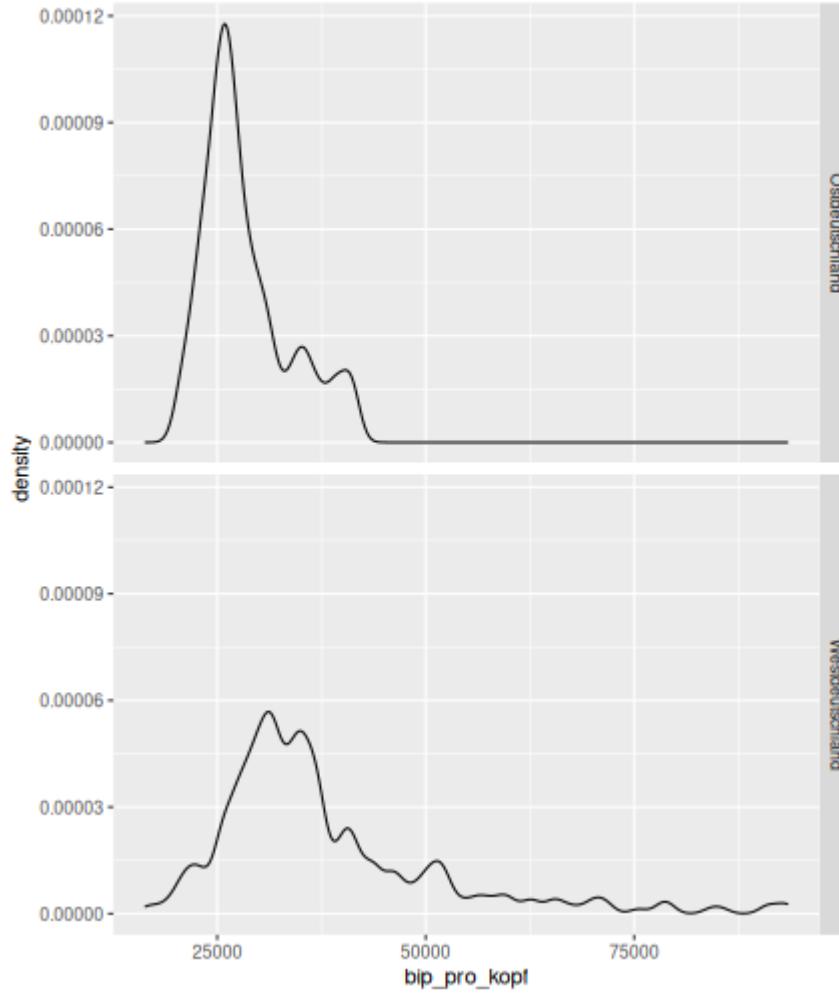
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_ko
```



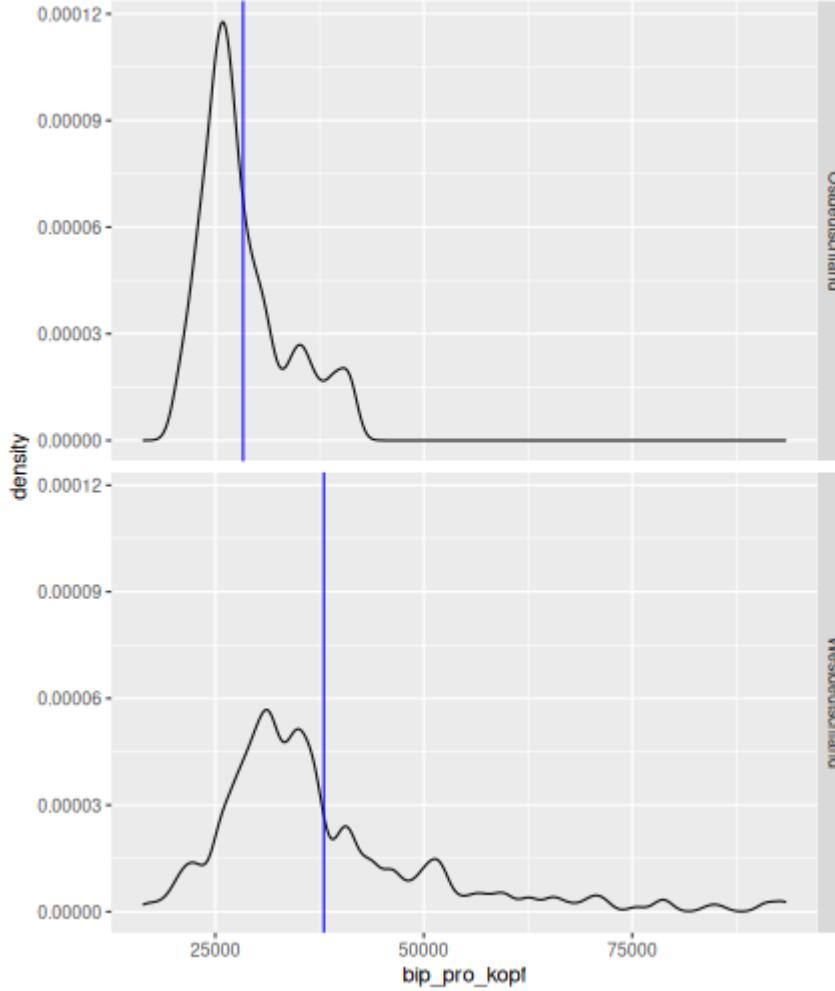
```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_ko
facet_grid(ost_name~.)
```



```
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_ko
facet_grid(ost_name~.) +
  geom_vline(data = mittel2, aes(xintercept = durchs
```

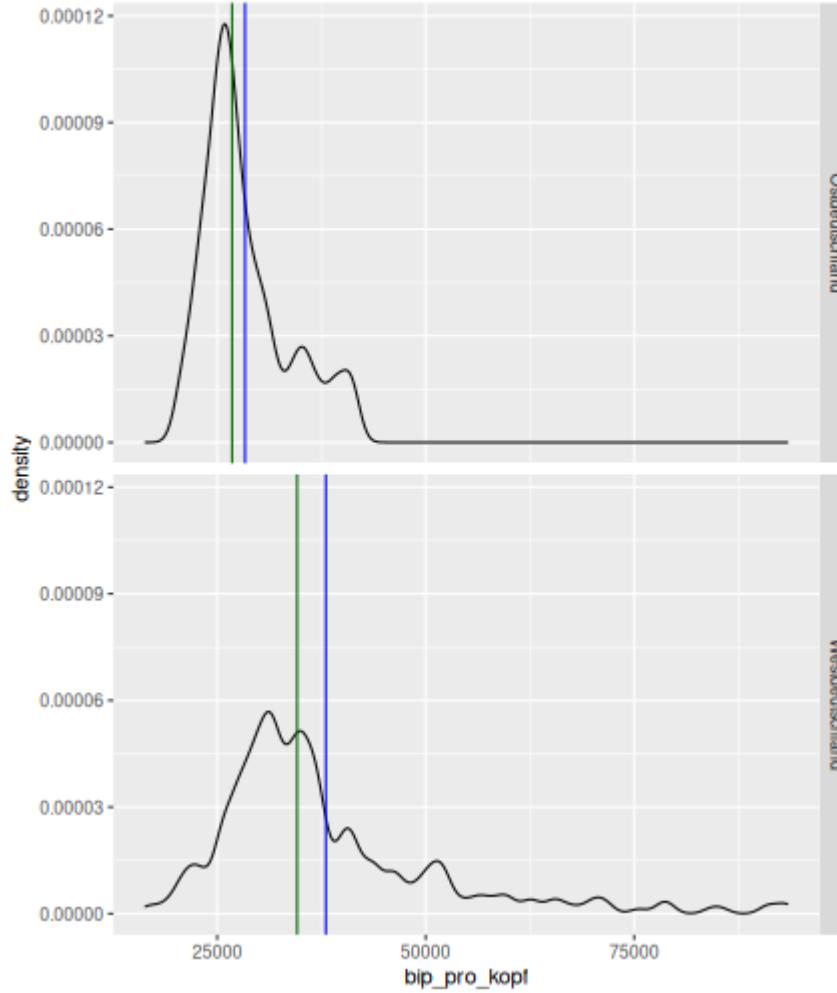


```

gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_kopf<100000))
  facet_grid(ost_name~.)
  geom_vline(data = mittel2, aes(xintercept = durchschnitt))
  geom_vline(data = mittel2, aes(xintercept = median))

```

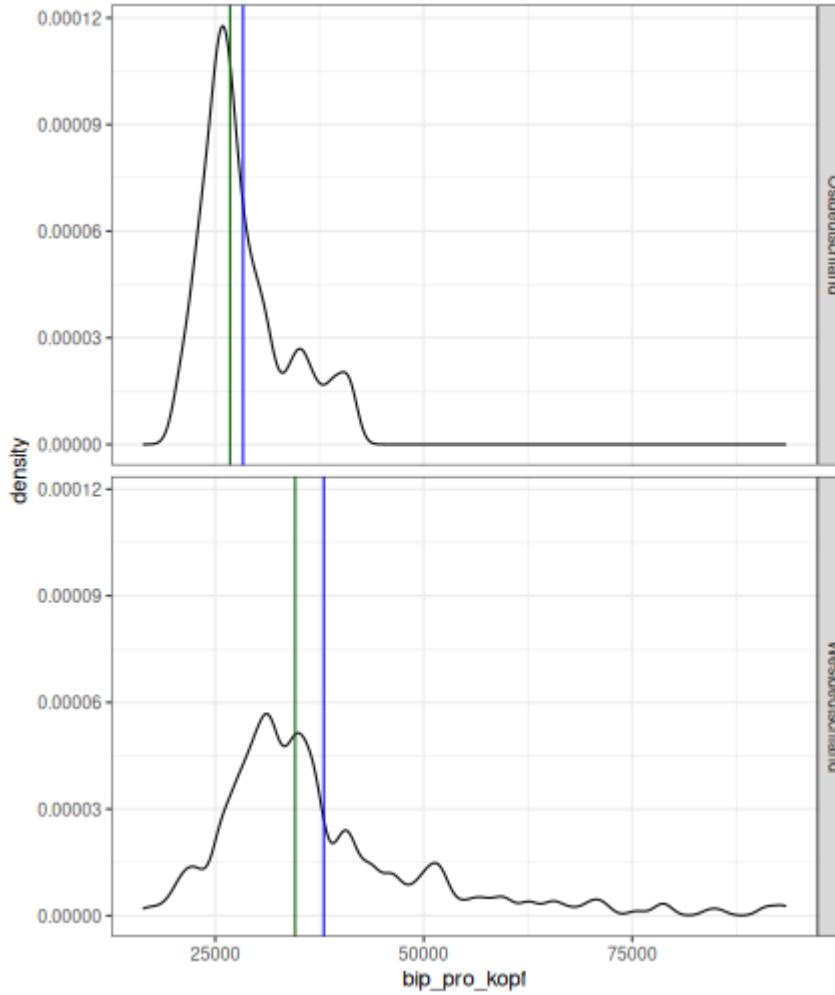


```

gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), media
ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_kopf<100000))
  facet_grid(ost_name~.)
  geom_vline(data = mittel2, aes(xintercept = durchschnitt))
  geom_vline(data = mittel2, aes(xintercept = median))
  theme_bw()

```

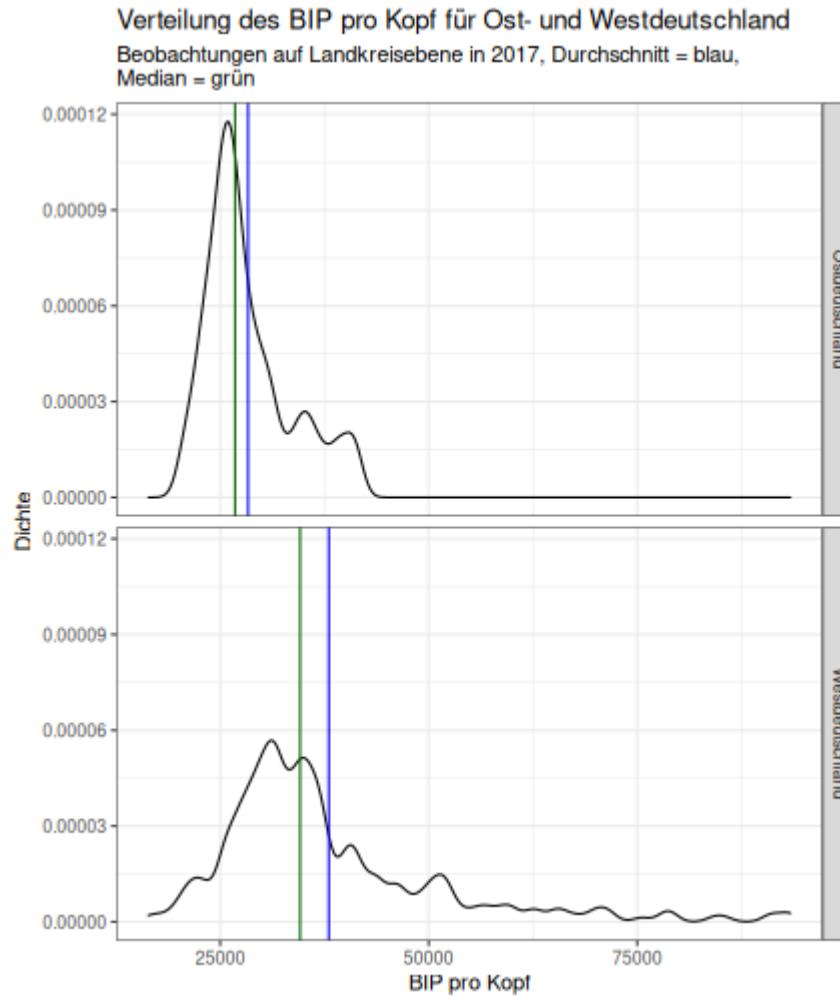


```

gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), median = median(bip_pro_kopf)) %>%
  ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_kopf < 100000)) +
  facet_grid(ost_name~.) +
  geom_vline(data = mittel2, aes(xintercept = durchschnitt), color = "blue") +
  geom_vline(data = mittel2, aes(xintercept = median), color = "green") +
  theme_bw() +
  labs(title = "Verteilung des BIP pro Kopf für Ost- und Westdeutschland",
       subtitle = "Beobachtungen auf Landkreisebene in 2017, Durchschnitt = blau, Median = grün",
       x = "BIP pro Kopf",
       y = "Dichte")

```

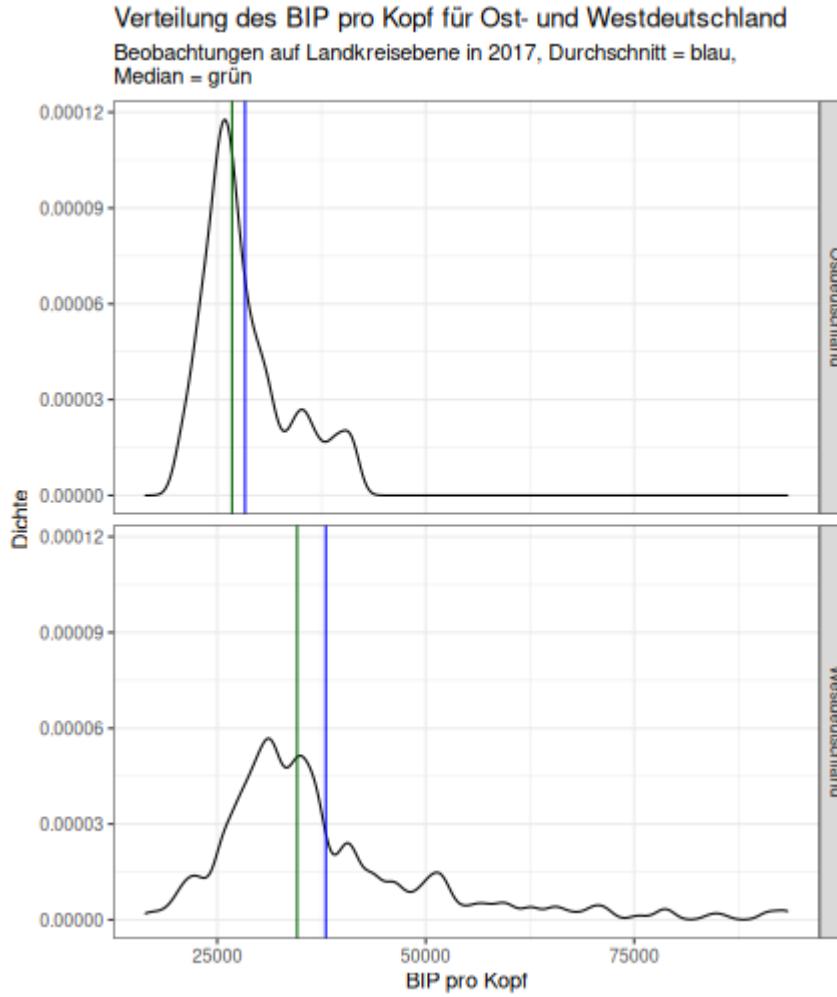


```

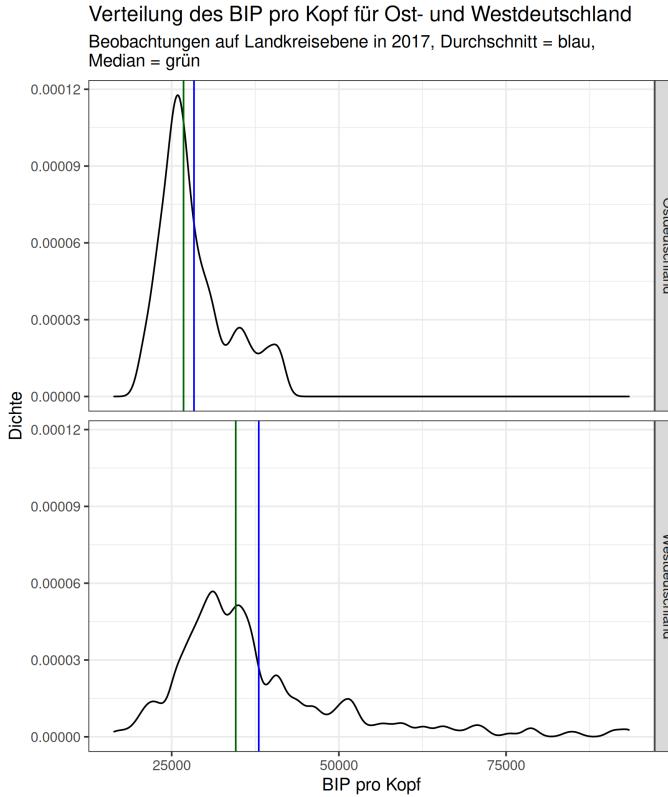
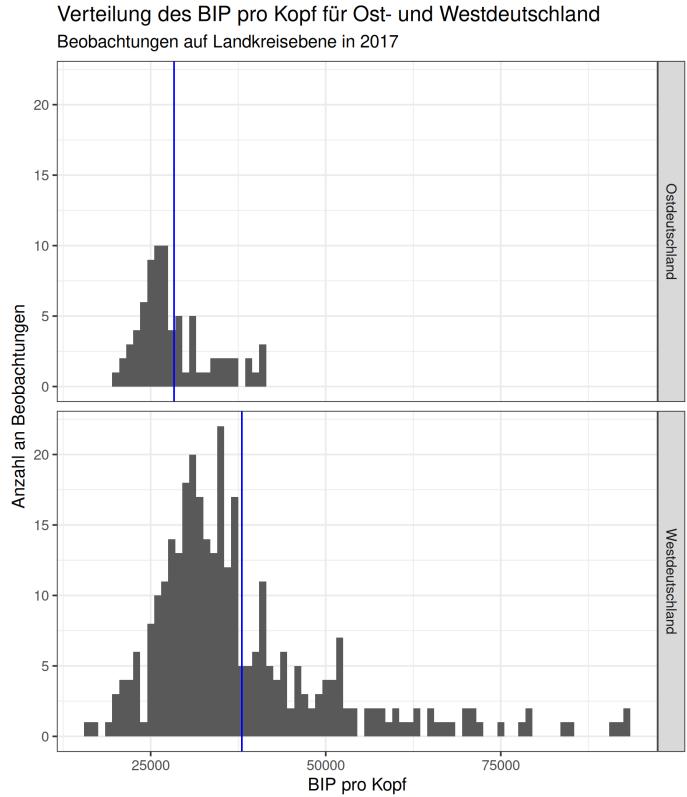
gesamtdaten %>%
  filter(bip_pro_kopf<100000) %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf), median = median(bip_pro_kopf)) %>%
  ungroup() -> mittel2

gesamtdaten %>%
  ggplot(aes(x = bip_pro_kopf)) +
  geom_density(data = filter(gesamtdaten, bip_pro_kopf < 100000)) +
  facet_grid(ost_name~.) +
  geom_vline(data = mittel2, aes(xintercept = durchschnitt), color = "blue") +
  geom_vline(data = mittel2, aes(xintercept = median), color = "green") +
  theme_bw() +
  labs(title = "Verteilung des BIP pro Kopf für Ost- und Westdeutschland",
       subtitle = "Beobachtungen auf Landkreisebene",
       x = "BIP pro Kopf",
       y = "Dichte")

```



Verteilung des BIP pro Kopf in 2017



Verteilung des BIP pro Kopf in 2017

Kendichteschätzer und Histogramm bestätigen das Bild des Boxplots:

- + Deutliche Unterschiede zwischen ost- und westdeutschend Landkreisen in 2017
- + Deutlich mehr Ausreißer nach oben bei westdeutschen Landkreisen
- + Verteilung ist für ostdeutsche Landkreise enger um den Mittelwert für das BIP pro Kopf von 28338€
- + Mittelwert und Median für westdeutsche Landkreise liegt deutlich weiter auseinander und zeigt, dass es hier mehr Ausreißer in den Daten gibt

Verschuldung der einzelnen Landkreise

Verschuldung

Warum könnte die Verschuldung des öffentlichen Haushalts ein Indikator für eine hohe Arbeitslosenquote sein?

Verschuldung

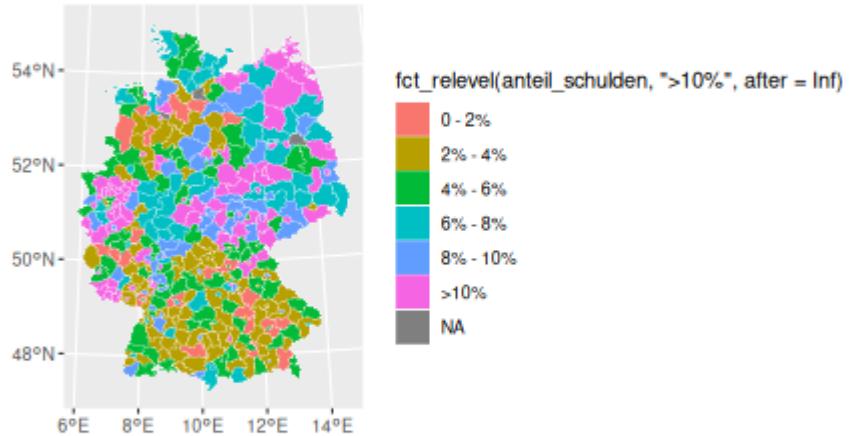
Warum könnte die Verschuldung des öffentlichen Haushalts ein Indikator für eine hohe Arbeitslosenquote sein?

Darstellung der Verschuldung der Landkreise mittels einer Deutschlandkarte.

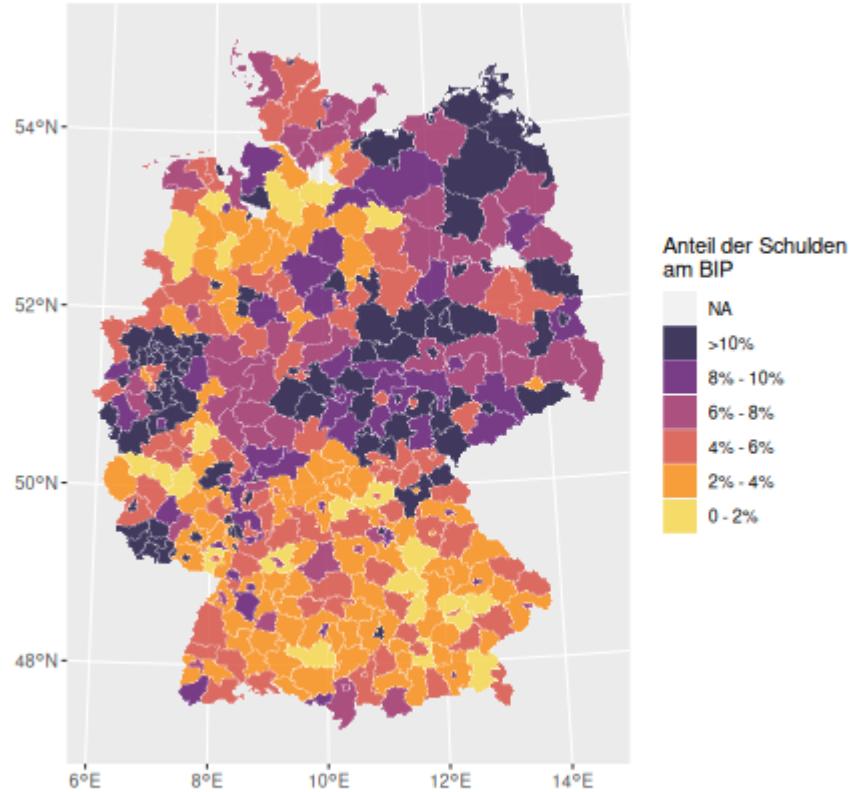
Beschreiben und interpretieren Sie die folgende Grafik.

```
ggplot(  
# define main data source  
data = schulden_landkreise_anteil  
)
```

```
ggplot(  
  # define main data source  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", after = Inf)  
    ),  
    color = "white",  
    size = 0.1  
)
```



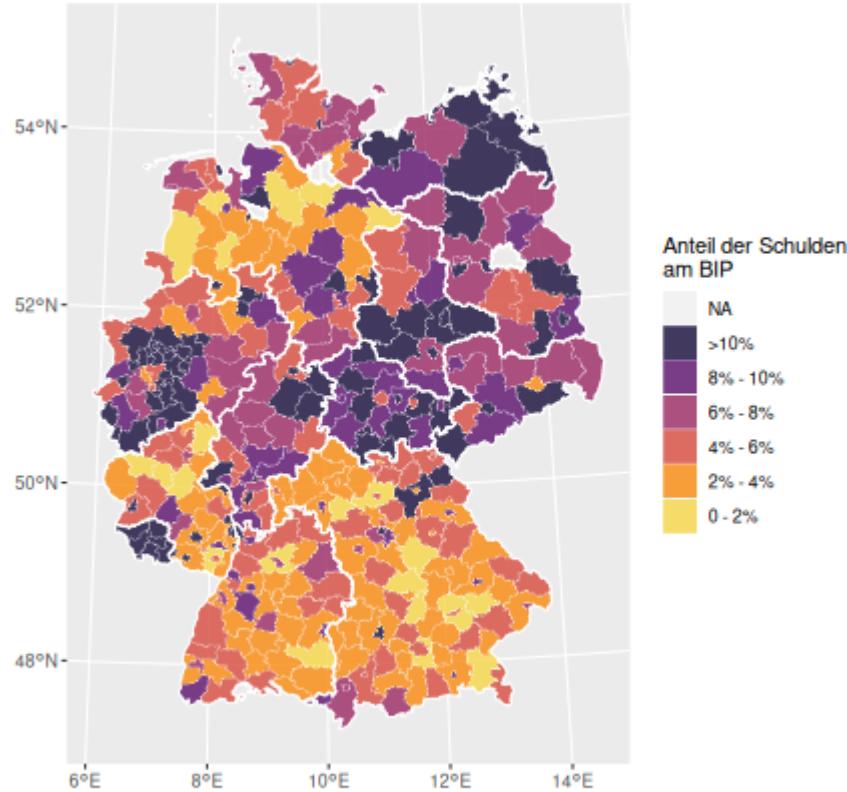
```
ggplot(  
  # define main data source  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af  
    ),  
    color = "white",  
    size = 0.1  
  ) +  
  # use the Viridis color scale  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "Anteil der Schulden\nam BIP",  
    alpha = 0.8, # make fill a bit brighter  
    begin = 0.1,  
    end = 0.9,  
    direction = -1,  
    guide = guide_legend(reverse = T))
```



```

ggplot(
# define main data source
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af
    ),
    color = "white",
    size = 0.1
  ) +
# use the Viridis color scale
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # make fill a bit brighter
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# use thicker white stroke for cantonal borders
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "white",
    size = 0.5
)

```

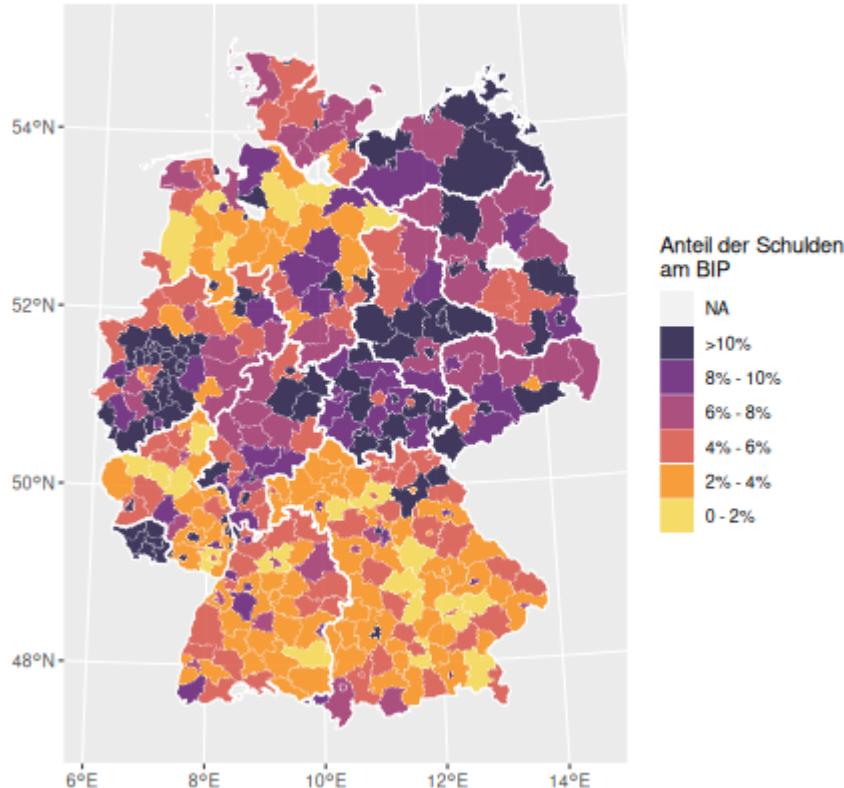


```

ggplot(
# define main data source
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af
    ),
    color = "white",
    size = 0.1
  ) +
# use the Viridis color scale
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # make fill a bit brighter
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# use thicker white stroke for cantonal borders
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "white",
    size = 0.5
  ) +
# add titles
  labs(x = NULL,
       y = NULL,
       title = "Wie verschuldet sind die deutschen Landkreise?",
       subtitle = "Öffentliche Schulden im Vergleich

```

Wie verschuldet sind die deutschen Landkreise?
Öffentliche Schulden im Vergleich zum BIP in 2017



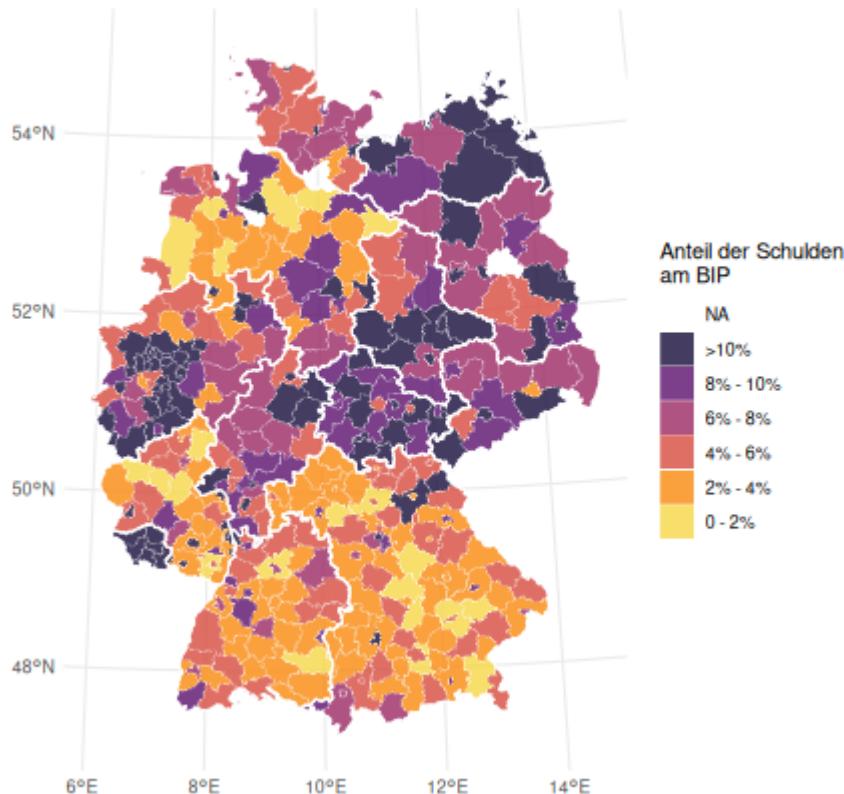
```
ggplot(  
# define main data source  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af  
    ),  
    color = "white",  
    size = 0.1  
  ) +  
  # use the Viridis color scale  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "Anteil der Schulden\nam BIP",  
    alpha = 0.8, # make fill a bit brighter  
    begin = 0.1,  
    end = 0.9,  
    direction = -1,  
    guide = guide_legend(reverse = T)) +  
  # use thicker white stroke for cantonal borders  
  geom_sf(  
    data = bundesland,  
    fill = "transparent",  
    color = "white",  
    size = 0.5  
  ) +  
  # add titles  
  labs(x = NULL,  
       y = NULL,  
       title = "Wie verschuldet sind die deutschen L  
       subtitle = "Öffentliche Schulden im Vergleich  
       theme_minimal() -> plot_schulden_lk
```

```

ggplot(
# define main data source
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af
    ),
    color = "white",
    size = 0.1
  ) +
# use the Viridis color scale
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # make fill a bit brighter
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# use thicker white stroke for cantonal borders
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "white",
    size = 0.5
  ) +
# add titles
  labs(x = NULL,
       y = NULL,
       title = "Wie verschuldet sind die deutschen Landkreise?",
       subtitle = "Öffentliche Schulden im Vergleich zum BIP in 2017",
       theme_minimal() -> plot_schulden_lk
)

```

Wie verschuldet sind die deutschen Landkreise?
Öffentliche Schulden im Vergleich zum BIP in 2017



Verschuldung

Beschreibung:

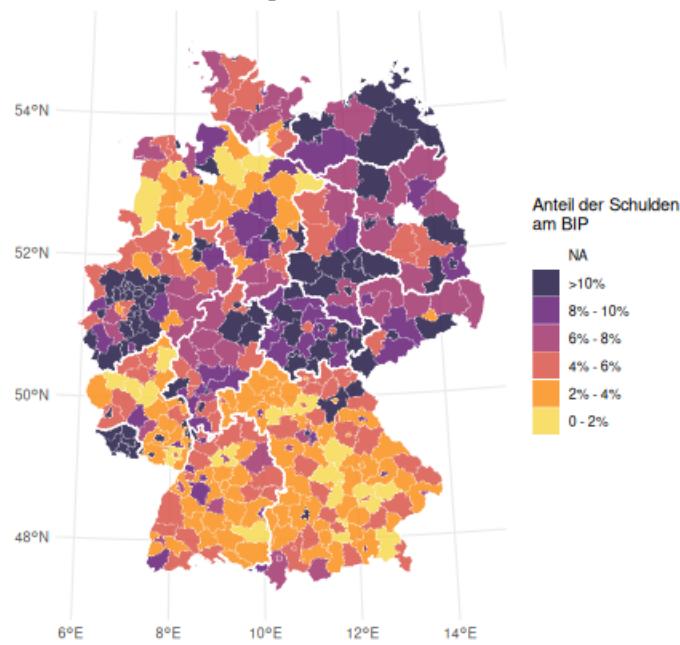
- + Niedrige Verschuldung im Verhältnis zum BIP: Bayern, Baden-Württemberg, Rheinland-Pflanz, Niedersachsen
- + Hohe Verschuldung: Nordrhein-Westfalen, Saarland, Sachse, Mecklenburg-Vorpommern
- + Mittlere Verschuldung: Brandenburg, Thüringen, Hessen

Interpretation:

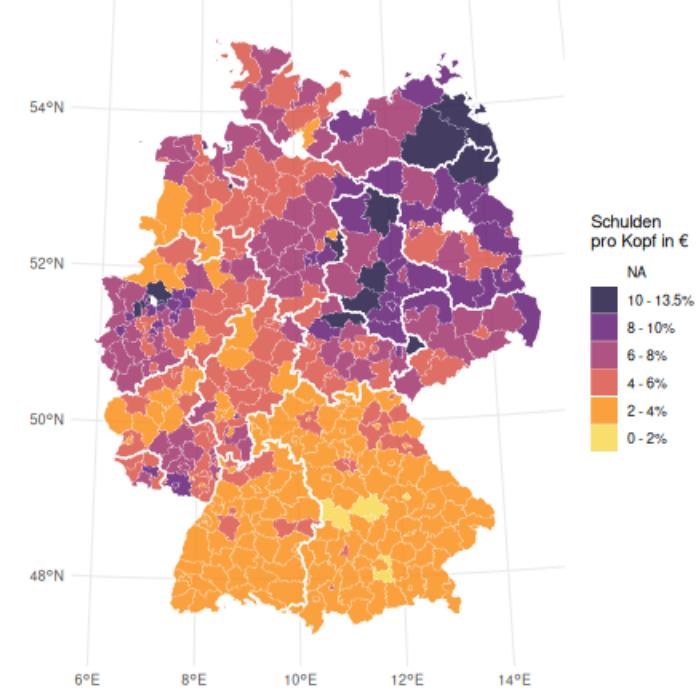
- + Strukturschwache Landkreise sind vermehrt in Ostdeutschland zu finden, allerdings scheint es eher ein Nord/Süd Gefälle als ein Ost/West Gefälle zu geben
- + Die ehemalige Herzammer der deutschen Industrie, das Ruhrgebiet, leidet unter dem Strukturwandel hin zu erneuerbaren Energien
 - + Es fallen hier wichtige Steuereinnahmen für die öffentliche Hand weg

Vergleich der Arbeitslosenquote und Verschuldung

Wie verschuldet sind die deutschen Landkreise?
Öffentliche Schulden im Vergleich zum BIP in 2017



Arbeitslosigkeit in Deutschland
Dargestellt ist die Arbeitslosenquote für alle Landkreise in 2017



Vergleich der Arbeitslosenquote und Verschuldung

- + Tendenziell sind die Landkreise mit höheren Schulden auch die mit einer höheren Arbeitslosenquote
- + Verschuldung könnte ein erklärender Faktor für die Arbeitslosenquote sein
- + Grafisch ist der Zusammenhang jedoch nicht eindeutig verifizierbar
 - + Um Zusammenhänge deutlich zu machen müssen wir uns der **bivariaten deskriptiven Statistik** bemühen, insbesondere **Streudiagrammen** und **Korrelationsmatrizen**

Karten sind eine schöne Art geografisch unterschiedliche Informationen darzustellen, allerdings ist das Auge schlecht darin Farbverläufe zu unterscheiden!

Bei Karten immer eine sehr kontrastreiche Farbpalette verwenden!

Bivariate deskriptive Analyse

Die Korrelation

Bisher: Univariate Analyse, d.h. nur eine Variable

Jetzt: Bivariate Analyse, d.h. Zusammenhang zwischen **zwei** Variablen untersuchen

Hierzu nutzen wir die Korrelation der Variablen!

Der Korrelationskoeffizient für zwei Variablen $(x_1, y_1), \dots, (x_n, y_n)$ ist definiert als:

$$\rho = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right)$$

mit μ_x, μ_y als Mittelwerte von x_1, \dots, x_n und y_1, \dots, y_n . σ_x, σ_y sind die Standardabweichungen von diesem Mittelwert.
 ρ wird üblicherweise genutzt um den Korrelationskoeffizienten zu bezeichnen.

Die Korrelation

Bisher: Univariate Analyse, d.h. nur eine Variable

Jetzt: Bivariate Analyse, d.h. Zusammenhang zwischen **zwei** Variablen untersuchen

Hierzu nutzen wir die Korrelation der Variablen!

Der Korrelationskoeffizient für zwei Variablen $(x_1, y_1), \dots, (x_n, y_n)$ ist definiert als:

$$\rho = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right)$$

mit μ_x, μ_y als Mittelwerte von x_1, \dots, x_n und y_1, \dots, y_n . σ_x, σ_y sind die Standardabweichungen von diesem Mittelwert. ρ wird üblicherweise genutzt um den Korrelationskoeffizienten zu bezeichnen.

Wie hängt die Arbeitslosenquote in den einzelnen Landkreisen mit deren BIP-Wachstum zusammen?

Korrelation zwischen Arbeitslosenquote und BIP-Wachstum

Wir können uns die oben beschriebene Formel bzgl. des Zusammenhangs von zwei Variablen immer auch grafisch verdeutlichen

- + Wir haben zwei Dimensionen
 - + Variable x: BIP-Wachstum
 - + Variable y: Arbeitslosenquote

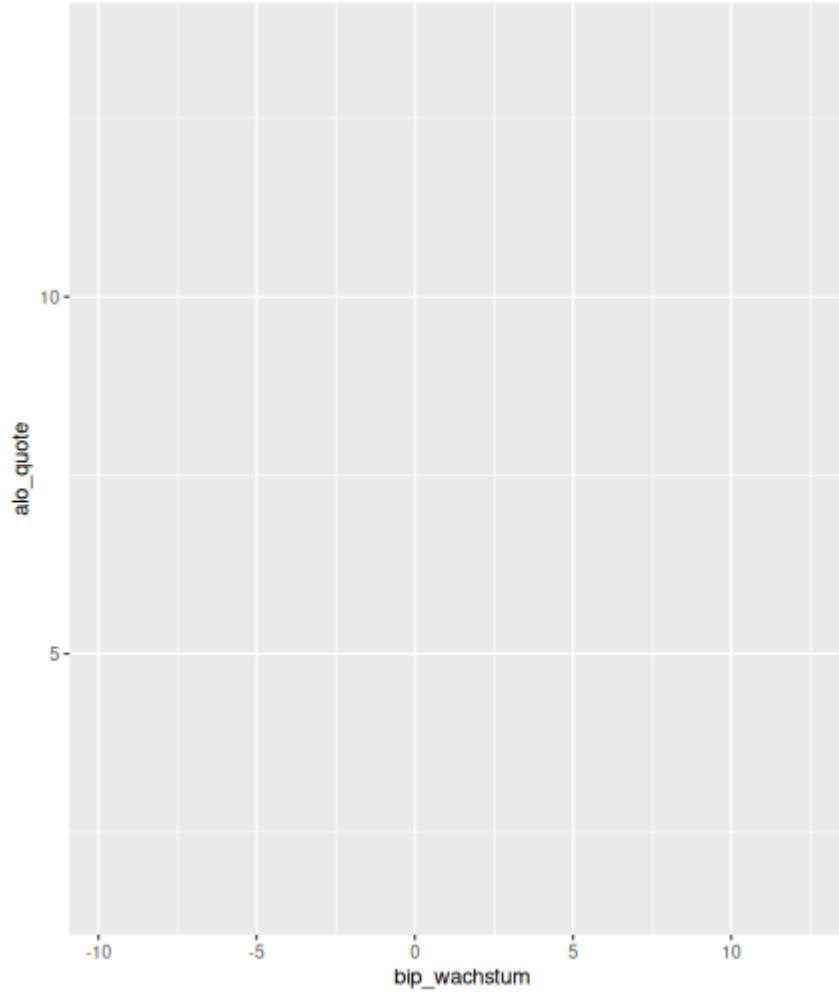
Im Streudiagramm können wir Variable x auf der x-Achse und Variable y auf der y-Achse abtragen

gesamtdaten

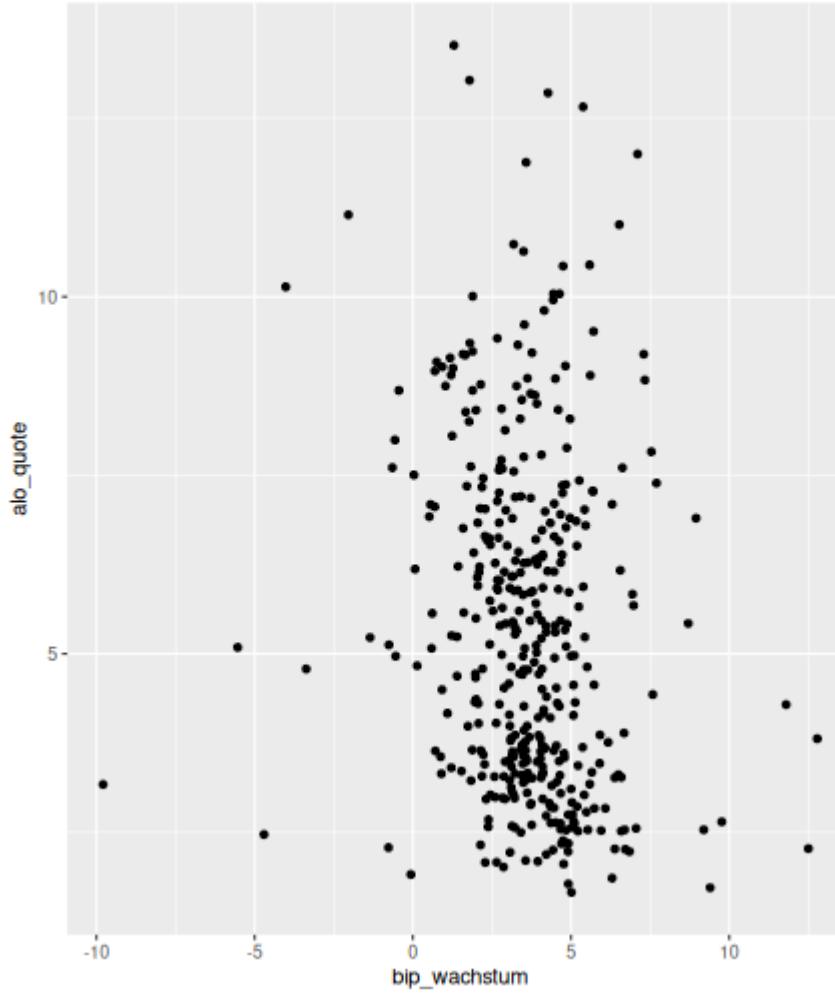
```
## # A tibble: 401 x 18
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name
##   <chr>           <dbl> <fct>          <chr>      <fct>
## 1 01001            4512 Flensburg    01        Schleswig-Hols...
## 2 01002            12345 Kiel         01        Schleswig-Hols...
## 3 01003            9692 Lübeck       01        Schleswig-Hols...
## 4 01004            3836 Neumünster   01        Schleswig-Hols...
## 5 01051            4632 Dithmarschen 01        Schleswig-Hols...
## 6 01053            5592 Herzogtum Lau... 01        Schleswig-Hols...
## 7 01054            5657 Nordfriesland 01        Schleswig-Hols...
## 8 01055            5748 Ostholstein   01        Schleswig-Hols...
## 9 01056            8599 Pinneberg    01        Schleswig-Hols...
## 10 01057           3264 Plön         01        Schleswig-Hols...
## # ... with 391 more rows, and 13 more variables: Schulden_pro_kopf_lk <dbl>,
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name.x <chr>, ost_name.y <chr>,
## #   bip_wachstum <dbl>, Jahr <dbl>, anteil_schulden <dbl>
```

```
gesamtdaten %>%
```

```
  ggplot(aes(x = bip_wachstum, y = alo_quote))
```

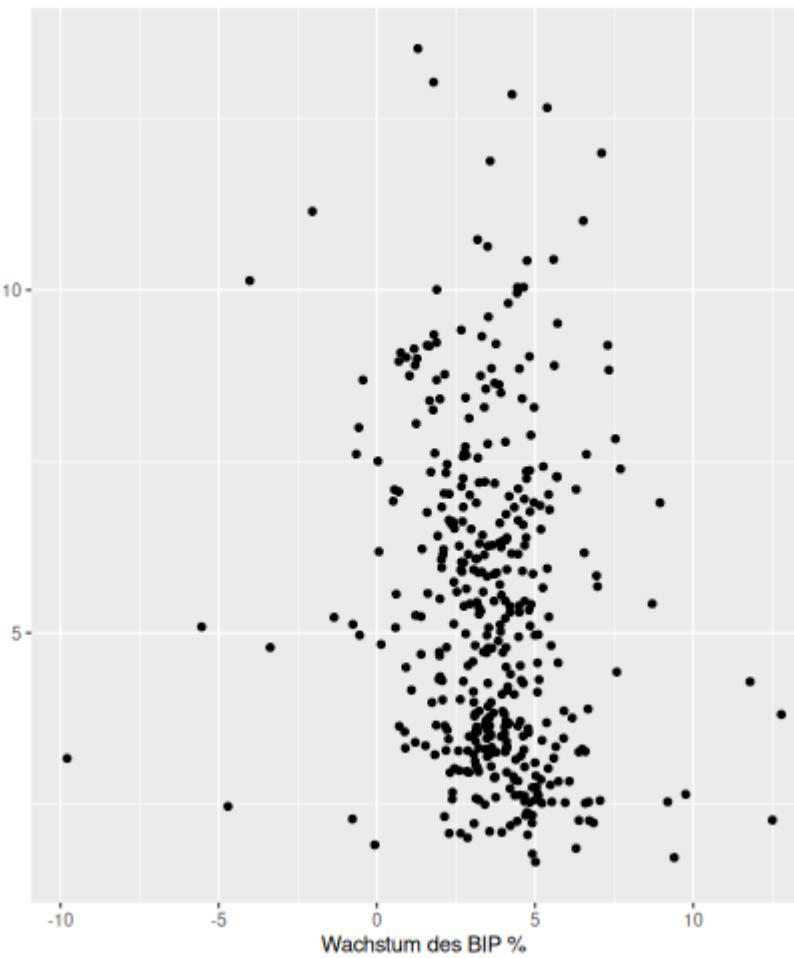


```
gesamtdaten %>%  
  ggplot(aes(x = bip_wachstum, y = alo_quote)) +  
  geom_point()
```



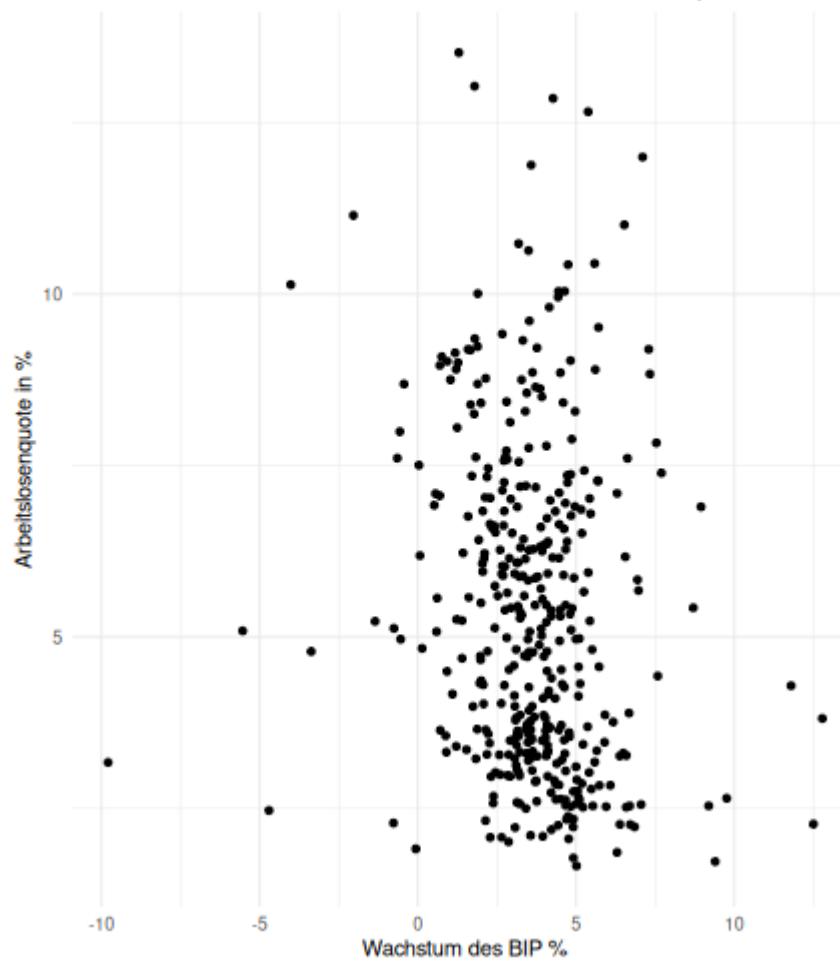
```
gesamtdaten %>%  
  ggplot(aes(x = bip_wachstum, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Wachstum des BIP %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation des BIP-Wachstums und d
```

Korrelation des BIP-Wachstums und der Arbeitslosenquote



```
gesamtdaten %>%
  ggplot(aes(x = bip_wachstum, y = alo_quote)) +
  geom_point() +
  labs( x = "Wachstum des BIP %",
        y = "Arbeitslosenquote in %",
        title = "Korrelation des BIP-Wachstums und d
  theme_minimal()
```

Korrelation des BIP-Wachstums und der Arbeitslosenquote



Korrelation zwischen Arbeitslosenquote und BIP-Wachstum

- + Es fallen die Ausreißer ins Auge (+10% und -10%)
 - + Vorheriges Jahr hohes/niedriges BIP, dadurch jetzt niedriges/hohes BIP-Wachstum
- + Insgesamt scheint der Zusammenhang jetzt nicht so stark zu sein
 - + Punktewolke deutet auf einen leicht negativen Zusammenhang hin

Korrelationskoeffizient:

```
cor(gesamtdaten$alo_quote,  
     gesamtdaten$bip_wachstum,  
     use = "pairwise.complete.obs")
```

```
## [1] -0.1521478
```

Korrelation zwischen Arbeitslosenquote und BIP-Wachstum

- + Es fallen die Ausreißer ins Auge (+10% und -10%)
 - + Vorheriges Jahr hohes/niedriges BIP, dadurch jetzt niedriges/hohes BIP-Wachstum
- + Insgesamt scheint der Zusammenhang jetzt nicht so stark zu sein
 - + Punktewolke deutet auf einen leicht negativen Zusammenhang hin

Korrelationskoeffizient:

```
cor(gesamtdaten$alo_quote,  
     gesamtdaten$bip_wachstum,  
     use = "pairwise.complete.obs")
```

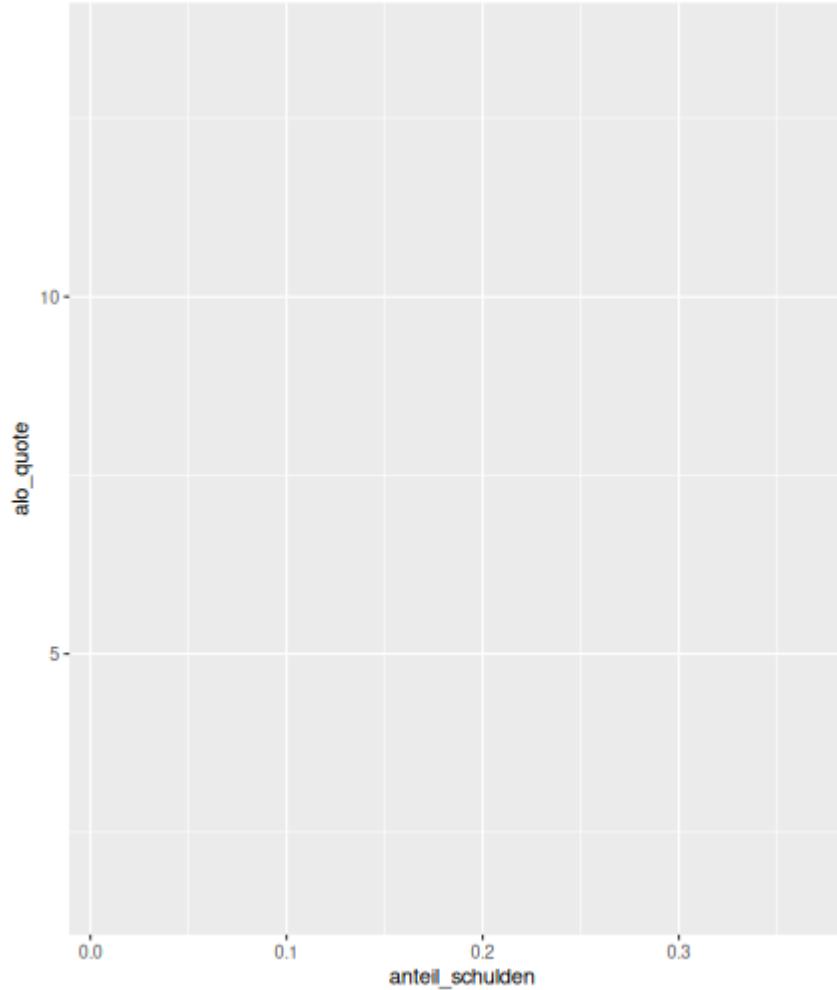
```
## [1] -0.1521478
```

Nun sollten wir noch die Korrelation zwischen Arbeitslosenquote und Verschuldung anschauen!

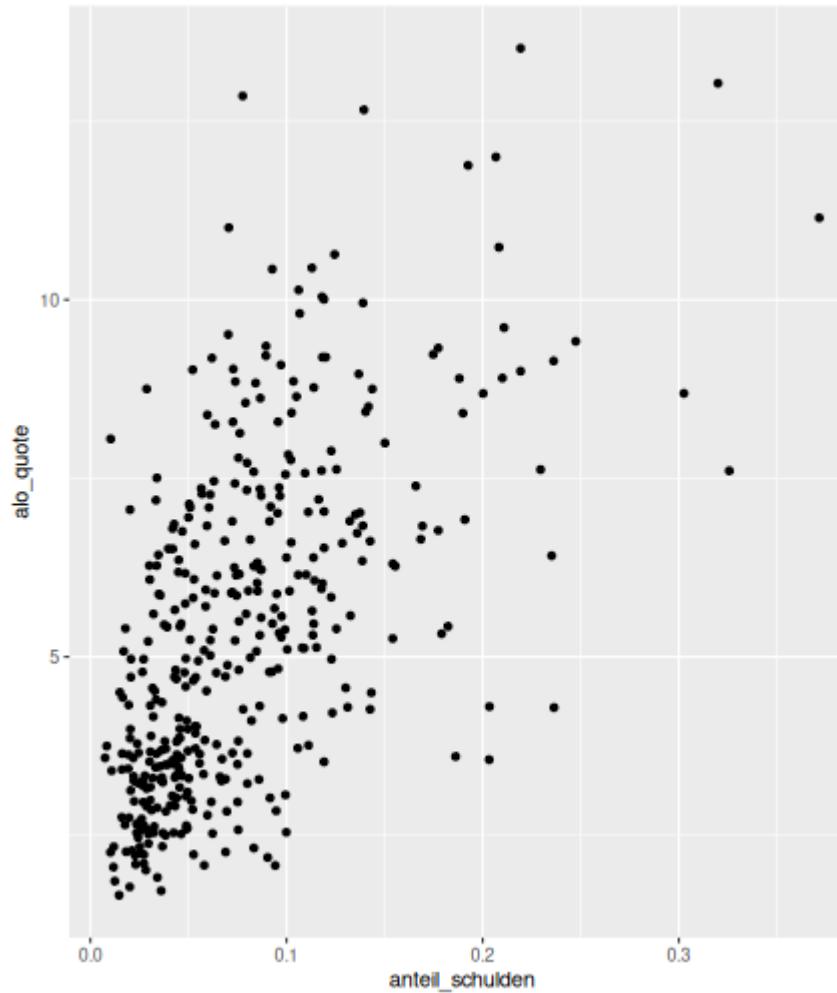
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g
```

```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten  
  
## # A tibble: 401 x 18  
##   Regionalschlues... total_alo landkreis_name bundesland bundesland_name  
##   <chr>              <dbl> <fct>      <chr>     <fct>  
## 1 01001               4512 Flensburg 01 Schleswig-Hols...  
## 2 01002               12345 Kiel    01 Schleswig-Hols...  
## 3 01003               9692 Lübeck  01 Schleswig-Hols...  
## 4 01004               3836 Neumünster 01 Schleswig-Hols...  
## 5 01051               4632 Dithmarschen 01 Schleswig-Hols...  
## 6 01053               5592 Herzogtum Lau... 01 Schleswig-Hols...  
## 7 01054               5657 Nordfriesland 01 Schleswig-Hols...  
## 8 01055               5748 Ostholstein 01 Schleswig-Hols...  
## 9 01056               8599 Pinneberg 01 Schleswig-Hols...  
## 10 01057              3264 Plön    01 Schleswig-Hols...  
## # ... with 391 more rows, and 13 more variables: Schulden_pro_kopf_lk <dbl>,  
## #   Einwohner <dbl>, Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>,  
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name.x <chr>, ost_name.y <chr>,  
## #   bip_wachstum <dbl>, Jahr <dbl>, anteil_schulden <dbl>
```

```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote))
```

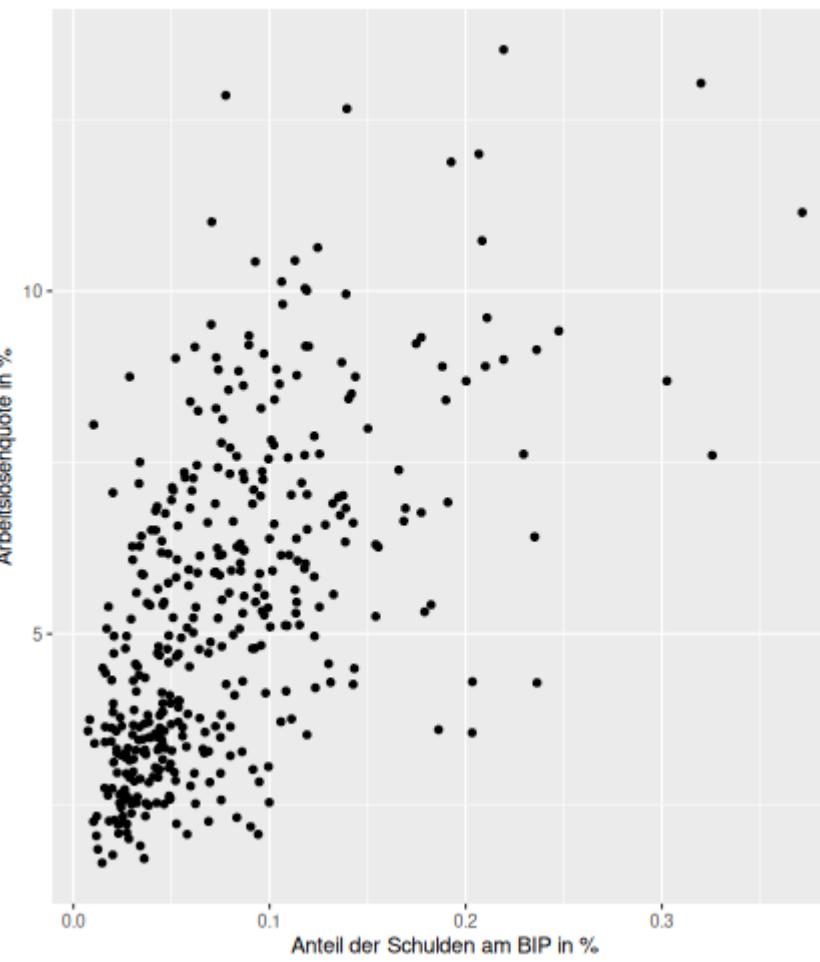


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point()
```

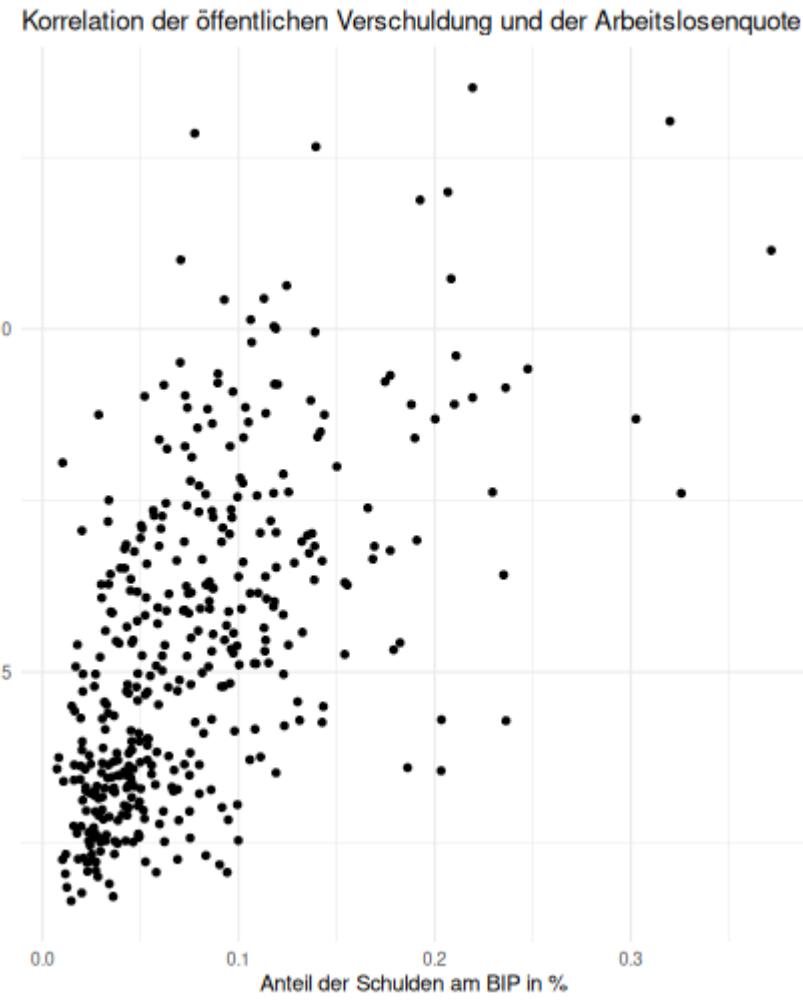


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Anteil der Schulden am BIP in %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation der öffentlichen Versch"
```

Korrelation der öffentlichen Verschuldung und der Arbeitslosenquote

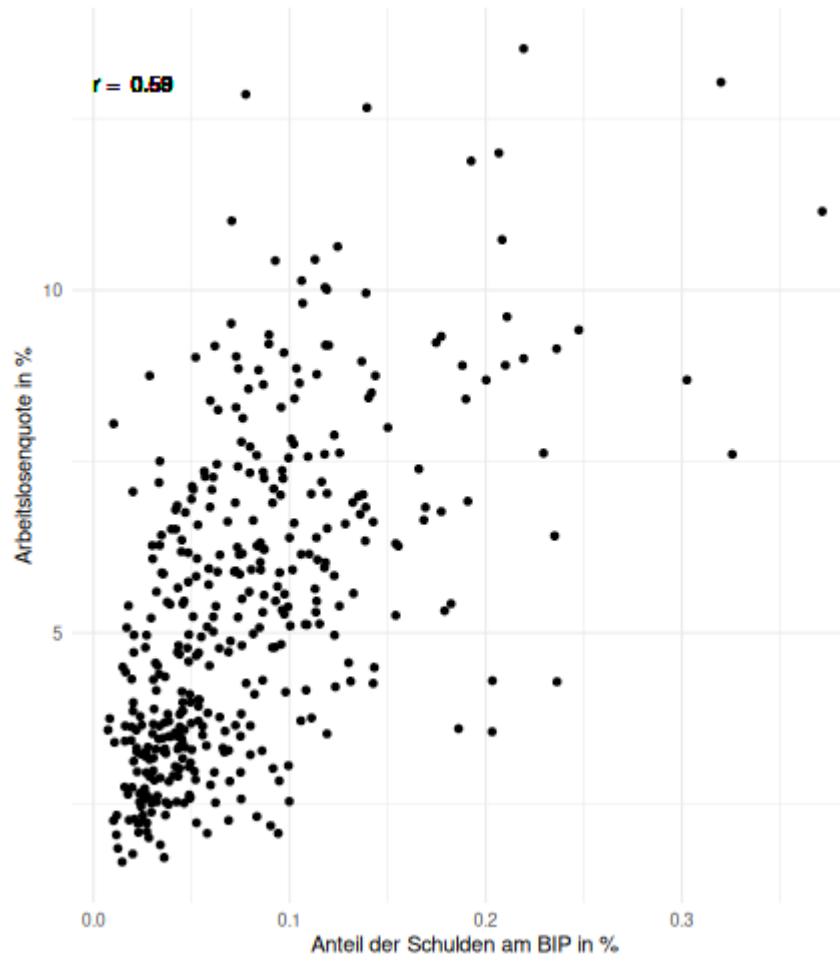


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Anteil der Schulden am BIP in %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation der öffentlichen Versch  
theme_minimal()
```



```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Anteil der Schulden am BIP in %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation der öffentlichen Versch  
theme_minimal() +  
  geom_text(x = 0.02, y =13, label = paste("r = ", a
```

Korrelation der öffentlichen Verschuldung und der Arbeitslosenquote



Korrelation zwischen Arbeitslosenquote und Verschuldung

Hier ist der positive Zusammenhang zwischen Verschuldung (x-Achse) und Arbeitslosenquote (y-Achse) deutlicher
Korrelationskoeffizient zeigt mit $\rho = 0.59$ auch einen starken Zusammenhang

ρ Beschreibung (nährungsweise)

- +/- 0.1-0.3 Schwacher
- +/- 0.3-0.5 Mittel
- +/- 0.5-0.8 Stark
- +/- 0.8-0.9 Sehr stark

Korrelation zwischen Arbeitslosenquote und Verschuldung

Hier ist der positive Zusammenhang zwischen Verschuldung (x-Achse) und Arbeitslosenquote (y-Achse) deutlicher
Korrelationskoeffizient zeigt mit $\rho = 0.59$ auch einen starken Zusammenhang

ρ Beschreibung (nährungsweise)

- +/- 0.1-0.3 Schwacher
- +/- 0.3-0.5 Mittel
- +/- 0.5-0.8 Stark
- +/- 0.8-0.9 Sehr stark

Wir sehen eine positive Korrelation zwischen der Verschuldung von Landkreisen und deren Arbeitslosenquoten.

Interpretation der Korrelation

- + Hat an sich keine intuitive quantitative Interpretation
- + Ist eine univariate Repräsentation des Zusammenhangs zweier Variablen
- + Kann dabei helfen stark korrelierte Variablen im Datensatz aufzuzeigen
 - + Dies ist für eine spätere lineare Regression wichtig
 - + Stichwort Multikollinearität

Im nächsten Semester beschäftigen wir uns mit der linearen Regression, hier können die Koeffizienten direkt interpretiert werden.

Zusammenfassung und Ausblick

Dieses Semester: Deskriptiven Statistik

Nächstes Semester: Induktive Statistik, insbesondere durch lineare Regressionen

Was haben wir bisher gelernt?

Zusammenfassung und Ausblick

Dieses Semester: Deskriptiven Statistik

Nächstes Semester: Induktive Statistik, insbesondere durch lineare Regressionen

Was haben wir bisher gelernt?

- ✚ Daten in R einlesen
- ✚ Diese Daten kompakt mittels Tabellen und Grafiken beschreiben
- ✚ Den Zusammenhang einzelner Variablen untersuchen

Übungsaufgaben

Im ersten Teil der Case Study hatten Sie sich noch die durchschnittlichen Einkommen auf Landkreisebene in R eingelesen. Nun sollten Sie diese Tabelle deskriptiv analysieren:

- ✚ Erstellen Sie eine deskriptive Tabelle, welche das Einkommen für das Jahr 2017 darstellt. Wie ist hier die Verteilung der Einkommen?
 - ✚ Beschreiben Sie Mittelwert, Standardabweichung, sowie Median
- ✚ Erstellen Sie ein Liniendiagramm zu der Entwicklung des Einkommensniveaus in den einzelnen Landkreisen seit 2000. Sie können sich hierbei an dem Diagramm zum BIP pro Kopf orientieren.
 - ✚ Hinweis: Mergen Sie zu dem Datensatz "Einkommen" zuerst noch die Information zu "Landkreis_name, Bundesland_name und ost_name" hinzu (siehe auch hierzu [diesen Abschnitt](#))
- ✚ Erstellen Sie eine Karte zum Einkommensniveau der einzelnen Landkreise. Sie können sich hierbei an der Karte zur Verschuldung orientieren.
- ✚ Erstellen Sie eine Korrelationstablle zwischen Arbeitslosenquote, Anteil Schulden, BIP pro Kopf und Einkommen. Sie können sich hierbei an der [Tabelle der Korrelationen aus diesem Abschnitt](#) orientieren.