

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!

Warum das Ganze?

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!

Warum das Ganze?

- + Durch die Deadlines sollten Sie sich mit dem Stoff auseinandersetzen
- + Github, R, RStudio und RMarkdown müssen Sie in den Projekten nutzen → Üben mit RTutor
- + Visualisierung, Interpretation und Präsentation in den Projekten gefragt → Üben mit der Case-Study

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

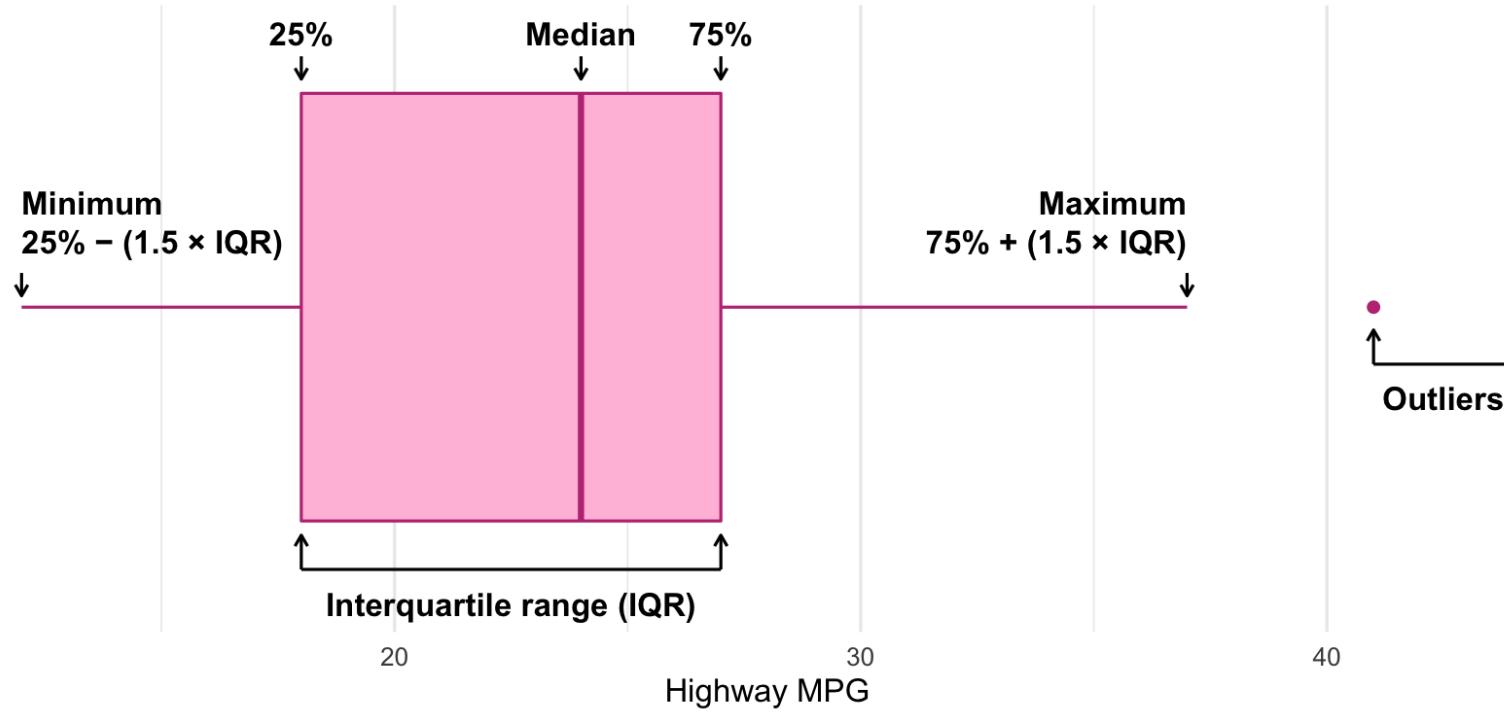
Univariate deskriptive Statistik

- + 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 × 2  
##   total_erw total_einwohner  
##       <dbl>          <dbl>  
## 1     41029891        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 × 2  
##   total_erw total_einwohner  
##       <dbl>          <dbl>  
## 1     41029891        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:

character 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: character

```
skim_variable  n_missing complete_rate min max empty n_unique whitespace
```

	n_missing	complete_rate	min	max	empty	n_unique	whitespace
landkreis_name	0	1	3	33	0	379	0

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

Woher kommt dies?

Anzahl an Beobachtungen

Variable type: character

```
skim_variable  n_missing complete_rate min max empty n_unique whitespace
```

	n_missing	complete_rate	min	max	empty	n_unique	whitespace
landkreis_name	0	1	3	33	0	379	0

- + 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	5.34	2.38	1.64	3.37	5.05	6.90	13.44
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	36827.71	16196.93	15737.94	27544.90	32592.40	39857.83	165520.71

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 × 1
##   landkreis_name
##   <chr>
## 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 × 1
##   landkreis_name
##   <chr>
## 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	5.34	2.38	1.64	3.37	5.05	6.90	13.44
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	36827.71	16196.93	15737.94	27544.90	32592.40	39857.83	165520.71

Beschreibung der Tabelle

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	5.34	2.38	1.64	3.37	5.05	6.90	13.44
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	36827.71	16196.93	15737.94	27544.90	32592.40	39857.83	165520.71

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	5.34	2.38	1.64	3.37	5.05	6.90	13.44
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	36827.71	16196.93	15737.94	27544.90	32592.40	39857.83	165520.71

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€

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

BIP pro Kopf

Mittelwert: 36828€

- + Insgesamt recht hoch
- + Starker Wirtschaftsstandort Deutschland

Standardabweichung: 16197€

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

Median: 32592€

- + 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 × 12
##   Regionalsch...¹ total...² landk...³ bunde...⁴ bunde...⁵ Schul...⁶ Einwo...⁷ Schul...⁸   lk
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb... 01     Schles...  5791.   87770  5.08e8 3.67e
## 2 01002           12345 Kiel     01     Schles...  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck   01     Schles...  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01     Schles...  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01     Schles...  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01     Schles...  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01     Schles...  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01     Schles...  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01     Schles...  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01     Schles...  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 3 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, and abbreviated variable names `¹Regionalschluessel`,
## #   `²total_alo`, `³landkreis_name`, `⁴bundesland`, `⁵bundesland_name`,
## #   `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

```
gesamtdaten %>%
  group_by( bundesland_name )
```

```
## # A tibble: 401 × 12
## # Groups:   bundesland_name [16]
##   Regionalsch...¹ total...² landk...³ bunde...⁴ bunde...⁵ Schul...⁶ Einwo...⁷ Schul...⁸   lk
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensbu... 01     Schles...  5791.   87770  5.08e8 3.67e
## 2 01002           12345 Kiel      01     Schles...  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck   01     Schles...  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01     Schles...  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01     Schles...  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01     Schles...  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01     Schles...  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01     Schles...  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinnebe... 01     Schles...  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön      01     Schles...  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 3 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, and abbreviated variable names ¹Regionalschlüssel,
## #   ²total_alo, ³landkreis_name, ⁴bundesland, ⁵bundesland_name,
## #   ⁶Schulden_pro_kopf_lk, ⁷Einwohner, ⁸Schulden_gesamt
```

```
gesamtdaten %>%
```

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

```
## # A tibble: 16 × 4
##   bundesland_name     mean_alo    sd_alo median_alo
##   <chr>          <dbl>      <dbl>      <dbl>
## 1 Baden-Württemberg  3.31      0.644     3.33
## 2 Bayern             3.04      0.770     3.00
## 3 Berlin              NA        NA        NA
## 4 Brandenburg         7.99      1.95      8.28
## 5 Bremen              8.95      2.54      8.95
## 6 Hamburg             NA        NA        NA
## 7 Hessen              5.01      1.35      5.06
## 8 Mecklenburg-Vorpommern  8.45      1.58      8.00
## 9 Niedersachsen       6.16      1.77      6.01
## 10 Nordrhein-Westfalen  7.15      2.43      6.88
## 11 Rheinland-Pfalz    5.31      1.47      5.26
## 12 Saarland            5.85      1.73      5.32
## 13 Sachsen             6.65      1.08      6.31
## 14 Sachsen-Anhalt      9.03      1.59      8.91
## 15 Schleswig-Holstein  6.36      1.00      6.82
## 16 Thüringen           6.30      1.80      5.70
```

```
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 × 4
##   bundesland_name      mean_alo   sd_alo   median
##   <chr>          <dbl>    <dbl>    <
## 1 Bayern           3.04     0.770
## 2 Baden-Württemberg 3.31     0.644
## 3 Hessen            5.01     1.35
## 4 Rheinland-Pfalz  5.31     1.47
## 5 Saarland          5.85     1.73
## 6 Niedersachsen    6.16     1.77
## 7 Thüringen         6.30     1.80
## 8 Schleswig-Holstein 6.36     1.00
## 9 Sachsen            6.65     1.08
## 10 Nordrhein-Westfalen 7.15     2.43
## 11 Brandenburg       7.99     1.95
## 12 Mecklenburg-Vorpommern 8.45     1.58
## 13 Bremen             8.95     2.54
## 14 Sachsen-Anhalt    9.03     1.59
```

Die Arbeitslosenquote auf Bundeslandebene

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

```
## # A tibble: 14 × 4
```

```
##   bунdesland_name    mean_alo   sd_alo median
##   <chr>            <dbl>     <dbl>   <dbl>
## 1 Bayern             3.04      0.770  
## 2 Baden-Württemberg 3.31      0.644  
## 3 Hessen             5.01      1.35   
## 4 Rheinland-Pfalz   5.31      1.47   
## 5 Saarland           5.85      1.73   
## 6 Niedersachsen     6.16      1.77   
## 7 Thüringen          6.30      1.80   
## 8 Schleswig-Holstein 6.36      1.00   
## 9 Sachsen            6.65      1.08   
## 10 Nordrhein-Westfalen 7.15      2.43  
## 11 Brandenburg       7.99      1.95  
## 12 Mecklenburg-Vorpommern 8.45      1.58  
## 13 Bremen             8.95      2.54  
## 14 Sachsen-Anhalt    9.03      1.59
```

Bundesland	Arbeitslosenquote		
	Mittelwert	Std.	Median
Bayern	3.04	0.77	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

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 × 4
##   bundesland_name     mean_alo   sd_alo median_alo
##   <chr>          <dbl>    <dbl>      <dbl>
## 1 Baden-Württemberg  3.31    0.644     3.33
## 2 Bayern            3.04    0.770     3.00
## 3 Berlin             NA     NA        NA
## 4 Brandenburg       7.99    1.95      8.28
## 5 Bremen             8.95    2.54      8.95
## 6 Hamburg            NA     NA        NA
## 7 Hessen              5.01    1.35      5.06
## 8 Mecklenburg-Vorpommern  8.45    1.58      8.00
## 9 Niedersachsen      6.16    1.77      6.01
## 10 Nordrhein-Westfalen 7.15    2.43      6.88
## 11 Rheinland-Pfalz   5.31    1.47      5.26
## 12 Saarland           5.85    1.73      5.32
## 13 Sachsen            6.65    1.08      6.31
## 14 Sachsen-Anhalt     9.03    1.59      8.91
## 15 Schleswig-Holstein 6.36    1.00      6.82
## 16 Thüringen          6.30    1.80      5.70
```

```
bula_data %>%
```

```
  arrange( mean_alo )
```

```
## # A tibble: 16 × 4
##   bundesland_name      mean_alo    sd_alo median_alo
##   <chr>          <dbl>     <dbl>      <dbl>
## 1 Bayern           3.04     0.770     3.00
## 2 Baden-Württemberg 3.31     0.644     3.33
## 3 Hessen            5.01     1.35      5.06
## 4 Rheinland-Pfalz  5.31     1.47      5.26
## 5 Saarland          5.85     1.73      5.32
## 6 Niedersachsen    6.16     1.77      6.01
## 7 Thüringen         6.30     1.80      5.70
## 8 Schleswig-Holstein 6.36     1.00      6.82
## 9 Sachsen            6.65     1.08      6.31
## 10 Nordrhein-Westfalen 7.15     2.43      6.88
## 11 Brandenburg       7.99     1.95      8.28
## 12 Mecklenburg-Vorpommern 8.45     1.58      8.00
## 13 Bremen             8.95     2.54      8.95
## 14 Sachsen-Anhalt    9.03     1.59      8.91
## 15 Berlin              NA       NA        NA
## 16 Hamburg             NA       NA        NA
```

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

```
## # A tibble: 14 × 4  
##   bundesland_name     mean_alo    sd_alo median_alo  
##   <chr>          <dbl>      <dbl>      <dbl>  
## 1 Bayern            3.04       0.770     3.00  
## 2 Baden-Württemberg 3.31       0.644     3.33  
## 3 Hessen             5.01       1.35      5.06  
## 4 Rheinland-Pfalz   5.31       1.47      5.26  
## 5 Saarland           5.85       1.73      5.32  
## 6 Niedersachsen     6.16       1.77      6.01  
## 7 Thüringen          6.30       1.80      5.70  
## 8 Schleswig-Holstein 6.36       1.00      6.82  
## 9 Sachsen            6.65       1.08      6.31  
## 10 Nordrhein-Westfalen 7.15       2.43      6.88  
## 11 Brandenburg       7.99       1.95      8.28  
## 12 Mecklenburg-Vorpommern 8.45       1.58      8.00  
## 13 Bremen             8.95       2.54      8.95  
## 14 Sachsen-Anhalt    9.03       1.59      8.91
```

```

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	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

```

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	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

```

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	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

```

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	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

```

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	3.00	
Baden-Württemberg	3.31	0.64	3.33	
Hessen	5.01	1.35	5.06	
Rheinland-Pfalz	5.31	1.47	5.26	
Saarland	5.85	1.73	5.32	
Niedersachsen	6.16	1.77	6.01	
Thüringen	6.30	1.80	5.70	
Schleswig-Holstein	6.36	1.00	6.82	
Sachsen	6.65	1.08	6.31	
Nordrhein-Westfalen	7.15	2.43	6.88	
Brandenburg	7.99	1.95	8.28	
Mecklenburg-Vorpommern	8.45	1.58	8.00	
Bremen	8.95	2.54	8.95	
Sachsen-Anhalt	9.03	1.59	8.91	

```

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	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

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	3.00
Baden-Württemberg	3.31	0.64	3.33
Hessen	5.01	1.35	5.06
Rheinland-Pfalz	5.31	1.47	5.26
Saarland	5.85	1.73	5.32
Niedersachsen	6.16	1.77	6.01
Thüringen	6.30	1.80	5.70
Schleswig-Holstein	6.36	1.00	6.82
Sachsen	6.65	1.08	6.31
Nordrhein-Westfalen	7.15	2.43	6.88
Brandenburg	7.99	1.95	8.28
Mecklenburg-Vorpommern	8.45	1.58	8.00
Bremen	8.95	2.54	8.95
Sachsen-Anhalt	9.03	1.59	8.91

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?

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

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine sehr niedrige Arbeitslosenquote (<4%)
- + Landkreise in den ostdeutschen Bundesländern leiden unter hohen Arbeitslosenquoten (>8%)

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine sehr niedrige Arbeitslosenquote (<4%)
- + Landkreise in den ostdeutschen Bundesländern leiden unter hohen Arbeitslosenquoten (>8%)
- + Standardabweichung bei allen Bundesländern vergleichbar
 - + Es gibt hier vermutlich keine großen Ausreißer bei den Arbeitslosenquoten in den Landkreisen

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

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

Sehr große Unterschiede in den durchschnittlichen Arbeitslosenquoten zwischen Landkreisen in 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 × 12
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002          12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck  01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01    Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01    Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01   Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01   Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01   Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01   Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01   Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 3 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, and abbreviated variable names `^Regionalschluessel',
## #   `^total_alo`, `^landkreis_name`, `^bundesland`, `^bundesland_name`,
## #   `^Schulden_pro_kopf_lk`, `^Einwohner`, `^Schulden_gesamt`
```

```
gesamtdaten %>%
```

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

```
## # A tibble: 401 x 14
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002           12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck   01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01     Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01     Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01     Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01     Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01     Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01     Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01     Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>,
## #   erw <dbl>, alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

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 × 14
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002          12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck  01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01    Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01    Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01   Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01   Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01   Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01   Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01   Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

```
gesamtdaten %>%
  group_by(ost_name)

## # A tibble: 401 × 14
## # Groups:   ost_name [2]
##   Regionalsch...¹ total...² landk...³ bunde...⁴ bunde...⁵ Schul...⁶ Einwo...⁷ Schul...⁸   lk
##   <chr>           <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001            4512 Flensb... 01     Schles...  5791.   87770  5.08e8 3.67e
## 2 01002            12345 Kiel     01     Schles...  3839.   247135 9.49e8 1.14e
## 3 01003            9692 Lübeck   01     Schles...  5567.   216739 1.21e9 9.16e
## 4 01004            3836 Neumün... 01     Schles...  5409.   78759  4.26e8 3.34e
## 5 01051            4632 Dithma... 01     Schles...  1670.   133684 2.23e8 4.47e
## 6 01053            5592 Herzog... 01     Schles...  1293.   195677 2.53e8 4.50e
## 7 01054            5657 Nordfr... 01     Schles...  2624.   165642 4.35e8 5.74e
## 8 01055            5748 Osthol... 01     Schles...  1890.   200931 3.80e8 5.27e
## 9 01056            8599 Pinneb... 01     Schles...  2225.   311713 6.94e8 9.07e
## 10 01057           3264 Plön     01     Schles...  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

```
gesamtdaten %>%  
  group_by(ost_name) %>%  
  summarise(mean_alo = mean(alo_quote, na.rm = T), s  
## # A tibble: 2 × 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    7.49    1.96    3.76    6.04    7.34    8.87    12.9  
## 2 Westdeutschland   4.84    2.18    1.64    3.21    4.25    6.22    13.4
```

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

```
## # A tibble: 2 × 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 7.49     1.96    3.76    6.04    7.34     8.87    12.9
## 2 Westdeutschland 4.84     2.18    1.64    3.21    4.25     6.22    13.4
```

```
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	7.49	1.96	3.76	6.04	7.34	8.87	12.90
Westdeutschland	4.84	2.18	1.64	3.21	4.25	6.22	13.44

```

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	7.49	1.96	3.76	6.04	7.34	8.87	12.90
Westdeutschland	4.84	2.18	1.64	3.21	4.25	6.22	13.44

```

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	7.49	1.96	3.76	6.04	7.34	8.87	12.90
Westdeutschland	4.84	2.18	1.64	3.21	4.25	6.22	13.44

```

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	7.49	1.96	3.76	6.04	7.34	8.87	12.90
Westdeutschland	4.84	2.18	1.64	3.21	4.25	6.22	13.44

```

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	7.49	1.96	3.76	6.04	7.34	8.87	12.90
Westdeutschland	4.84	2.18	1.64	3.21	4.25	6.22	13.44

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 Landkreisen der westdeutschen Bundesländer
- ✚ Die Standardabweichung ist vergleichbar
- ✚ Der Median liegt in den westdeutschen Landkreisen sogar noch deutlicher unter dem Mittelwert als in ostdeutschen
- ✚ Im **25% Quantil** in den **ostdeutschen Landkreisen** ist die Arbeitslosenquote bei **6,04%**
- ✚ Bei den **westdeutschen Landkreisen** ist das **75% Quantil** bei einer Arbeitslosenquote von **6,22%**!

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 × 14
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002          12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck  01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01    Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01    Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01   Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01   Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01   Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01   Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01   Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt`
```

```
gesamtdaten %>%
  group_by( bundesland_name )
```

```
## # A tibble: 401 × 14
## # Groups:   bundesland_name [16]
##   Regionalsch...¹ total...² landk...³ bunde...⁴ bunde...⁵ Schul...⁶ Einwo...⁷ Schul...⁸   lk
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensbu... 01     Schles...  5791.   87770  5.08e8 3.67e
## 2 01002           12345 Kiel      01     Schles...  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck   01     Schles...  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01     Schles...  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01     Schles...  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01     Schles...  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01     Schles...  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01     Schles...  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinnebe... 01     Schles...  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön      01     Schles...  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt`
```

```
gesamtdaten %>%
```

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

```
## # A tibble: 16 × 7  
##   bundesland_name  
##   <chr>  
## 1 Baden-Württemberg  
## 2 Bayern  
## 3 Berlin  
## 4 Brandenburg  
## 5 Bremen  
## 6 Hamburg  
## 7 Hessen  
## 8 Mecklenburg-Vorpommern  
## 9 Niedersachsen  
## 10 Nordrhein-Westfalen  
## 11 Rheinland-Pfalz  
## 12 Saarland  
## 13 Sachsen  
## 14 Sachsen-Anhalt  
## 15 Schleswig-Holstein  
## 16 Thüringen  
## # ... with abbreviated variable names  
## #   `¹sd_bip_kopf`, `²mean_schulden_kopf`,  
## #   `³sd_schulden`  
## #   `mean_alo` <dbl> `sd_alo` <dbl> `mean_bip_kopf` <dbl> `sd_bip...¹` <dbl> `mean_...²` <dbl> `sd_schulden` <dbl>
```

bundesland_name	mean_alo	sd_alo	mean_bip_kopf	sd_bip...¹	mean_...²	sd_schulden
Baden-Württemberg	3.31	0.644	43674.	12935.	2185.	161
Bayern	3.04	0.770	42891.	20008.	1897.	140
Berlin	NA	NA	NA	NA	NA	NA
Brandenburg	7.99	1.95	28803.	5854.	2831.	147
Bremen	8.95	2.54	42780.	10291.	NA	NA
Hamburg	NA	NA	NA	NA	NA	NA
Hessen	5.01	1.35	40083.	16756.	3731.	287
Mecklenburg-Vorpommern	8.45	1.58	28194.	6239.	3565.	166
Niedersachsen	6.16	1.77	35274.	22952.	1941.	138
Nordrhein-Westfalen	7.15	2.43	35808.	10895.	4243.	248
Rheinland-Pfalz	5.31	1.47	33954.	13965.	3131.	359
Saarland	5.85	1.73	32923.	7384.	5959.	136
Sachsen	6.65	1.08	28562.	5243.	2306.	67
Sachsen-Anhalt	9.03	1.59	26810.	4097.	2809.	109
Schleswig-Holstein	6.36	1.00	32494.	7691.	2602.	170
Thüringen	6.30	1.80	28009.	5926.	2832.	54

```
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 × 7

bundesland_name	mean_alo	sd_alo	mean_bip_kopf	sd_bip... ¹	mean_... ²	sd_sc...
Baden-Württemberg	3.31	0.644	43674.	12935.	2185.	161
Bayern	3.04	0.770	42891.	20008.	1897.	140
Berlin	NA	NA	NA	NA	NA	NA
Brandenburg	7.99	1.95	28803.	5854.	2831.	147
Bremen	8.95	2.54	42780.	10291.	NA	NA
Hamburg	NA	NA	NA	NA	NA	NA
Hessen	5.01	1.35	40083.	16756.	3731.	281
Mecklenburg-Vorpommern	8.45	1.58	28194.	6239.	3565.	166
Niedersachsen	6.16	1.77	35274.	22952.	1941.	138
Nordrhein-Westfalen	7.15	2.43	35808.	10895.	4243.	248
Rheinland-Pfalz	5.31	1.47	33954.	13965.	3131.	359
Saarland	5.85	1.73	32923.	7384.	5959.	136
Sachsen	6.65	1.08	28562.	5243.	2306.	67
Sachsen-Anhalt	9.03	1.59	26810.	4097.	2809.	109
Schleswig-Holstein	6.36	1.00	32494.	7691.	2602.	170
Thüringen	6.30	1.80	28009.	5926.	2832.	54

... with abbreviated variable names ¹sd_bip_kopf, ²mean_schulden_kopf,
³sd_schulden

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

bula_data_all %>%
  arrange( mean_alo )
```

```

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) )

```

	bundesland_name	mean_alo	sd_alo	mean_bip_kopf	sd_bip...¹	mean_...²	sd_s...
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	Bayern	3.04	0.770	42891.	20008.	1897.	140
## 2	Baden-Württemberg	3.31	0.644	43674.	12935.	2185.	161
## 3	Hessen	5.01	1.35	40083.	16756.	3731.	287
## 4	Rheinland-Pfalz	5.31	1.47	33954.	13965.	3131.	359
## 5	Saarland	5.85	1.73	32923.	7384.	5959.	136
## 6	Niedersachsen	6.16	1.77	35274.	22952.	1941.	138
## 7	Thüringen	6.30	1.80	28009.	5926.	2832.	54
## 8	Schleswig-Holstein	6.36	1.00	32494.	7691.	2602.	170
## 9	Sachsen	6.65	1.08	28562.	5243.	2306.	67
## 10	Nordrhein-Westfalen	7.15	2.43	35808.	10895.	4243.	248
## 11	Brandenburg	7.99	1.95	28803.	5854.	2831.	147
## 12	Mecklenburg-Vorpommern	8.45	1.58	28194.	6239.	3565.	166
## 13	Sachsen-Anhalt	9.03	1.59	26810.	4097.	2809.	109
## #	... with abbreviated variable names	¹sd_bip_kopf,	²mean_schulden_kopf,				
## #	³sd_schulden						

```

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	42891.12	20008.07	1896.94
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49
Hessen	5.01	1.35	40082.52	16755.57	3730.85
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70
Saarland	5.85	1.73	32923.33	7383.88	5958.80
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34
Thüringen	6.30	1.80	28009.18	5925.85	2832.09
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65
Sachsen	6.65	1.08	28562.48	5242.58	2305.58
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	2809.25

```

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	42891.12	20008.07	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49	1612.86
Hessen	5.01	1.35	40082.52	16755.57	3730.85	2870.38
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70	3593.36
Saarland	5.85	1.73	32923.33	7383.88	5958.80	1363.06
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34	1381.10
Thüringen	6.30	1.80	28009.18	5925.85	2832.09	547.15
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65	1700.98
Sachsen	6.65	1.08	28562.48	5242.58	2305.58	674.06
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47	2483.34
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16	1473.58
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62	1667.49
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	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	42891.12	20008.07	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49	1612.86
Hessen	5.01	1.35	40082.52	16755.57	3730.85	2870.38
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70	3593.36
Saarland	5.85	1.73	32923.33	7383.88	5958.80	1363.06
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34	1381.10
Thüringen	6.30	1.80	28009.18	5925.85	2832.09	547.15
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65	1700.98
Sachsen	6.65	1.08	28562.48	5242.58	2305.58	674.06
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47	2483.34
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16	1473.58
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62	1667.49
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	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	42891.12	20008.07	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49	1612.86
Hessen	5.01	1.35	40082.52	16755.57	3730.85	2870.38
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70	3593.36
Saarland	5.85	1.73	32923.33	7383.88	5958.80	1363.06
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34	1381.10
Thüringen	6.30	1.80	28009.18	5925.85	2832.09	547.15
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65	1700.98
Sachsen	6.65	1.08	28562.48	5242.58	2305.58	674.06
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47	2483.34
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16	1473.58
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62	1667.49
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	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	42891.12	20008.07	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49	1612.86
Hessen	5.01	1.35	40082.52	16755.57	3730.85	2870.38
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70	3593.36
Saarland	5.85	1.73	32923.33	7383.88	5958.80	1363.06
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34	1381.10
Thüringen	6.30	1.80	28009.18	5925.85	2832.09	547.15
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65	1700.98
Sachsen	6.65	1.08	28562.48	5242.58	2305.58	674.06
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47	2483.34
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16	1473.58
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62	1667.49
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	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	42891.12	20008.07	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49	1612.86
Hessen	5.01	1.35	40082.52	16755.57	3730.85	2870.38
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70	3593.36
Saarland	5.85	1.73	32923.33	7383.88	5958.80	1363.06
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34	1381.10
Thüringen	6.30	1.80	28009.18	5925.85	2832.09	547.15
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65	1700.98
Sachsen	6.65	1.08	28562.48	5242.58	2305.58	674.06
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47	2483.34
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16	1473.58
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62	1667.49
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	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	42891.12	20008.07	1896.94	1399.73
Baden-Württemberg	3.31	0.64	43673.84	12934.97	2185.49	1612.86
Hessen	5.01	1.35	40082.52	16755.57	3730.85	2870.38
Rheinland-Pfalz	5.31	1.47	33954.05	13965.48	3130.70	3593.36
Saarland	5.85	1.73	32923.33	7383.88	5958.80	1363.06
Niedersachsen	6.16	1.77	35274.47	22951.65	1941.34	1381.10
Thüringen	6.30	1.80	28009.18	5925.85	2832.09	547.15
Schleswig-Holstein	6.36	1.00	32494.36	7691.09	2601.65	1700.98
Sachsen	6.65	1.08	28562.48	5242.58	2305.58	674.06
Nordrhein-Westfalen	7.15	2.43	35807.86	10894.53	4243.47	2483.34
Brandenburg	7.99	1.95	28803.45	5853.89	2831.16	1473.58
Mecklenburg-Vorpommern	8.45	1.58	28193.92	6239.42	3564.62	1667.49
Sachsen-Anhalt	9.03	1.59	26810.40	4096.90	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

- ✚ Landkreise in Bundesländer mit niedrigen Arbeitslosenquoten haben durchschnittlich ein hohes BIP pro Kopf
- ✚ Ostdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf < 30000€
- ✚ Westdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf > 30000€
- ✚ Kein klares Bild der Landkreise hinsichtlich der Schulden pro Kopf

Arbeitslosenquote, BIP pro Kopf und Schulden pro Kopf

- ✚ Landkreise in Bundesländer mit niedrigen Arbeitslosenquoten haben durchschnittlich ein hohes BIP pro Kopf
- ✚ Ostdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf $< 30000\text{€}$
- ✚ Westdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf $> 30000\text{€}$
- ✚ Kein klares Bild der Landkreise hinsichtlich der 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 × 14
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002          12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck   01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01     Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01     Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01     Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01     Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01     Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01     Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01     Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

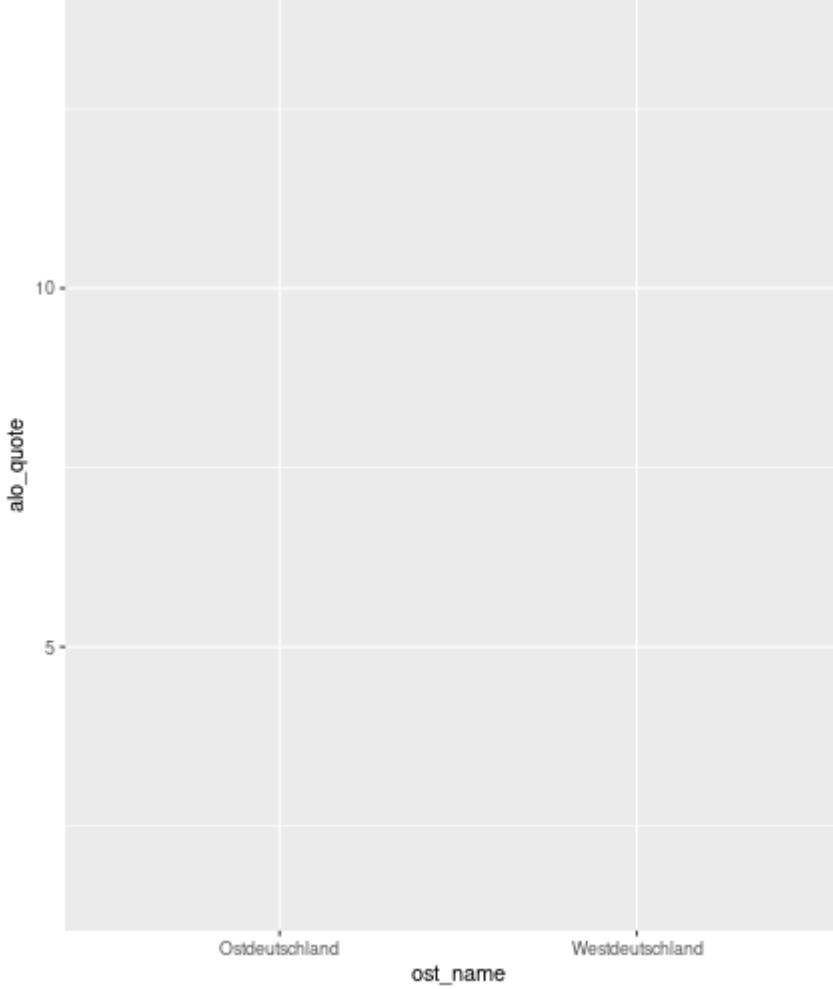
```
gesamtdaten %>%
```

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

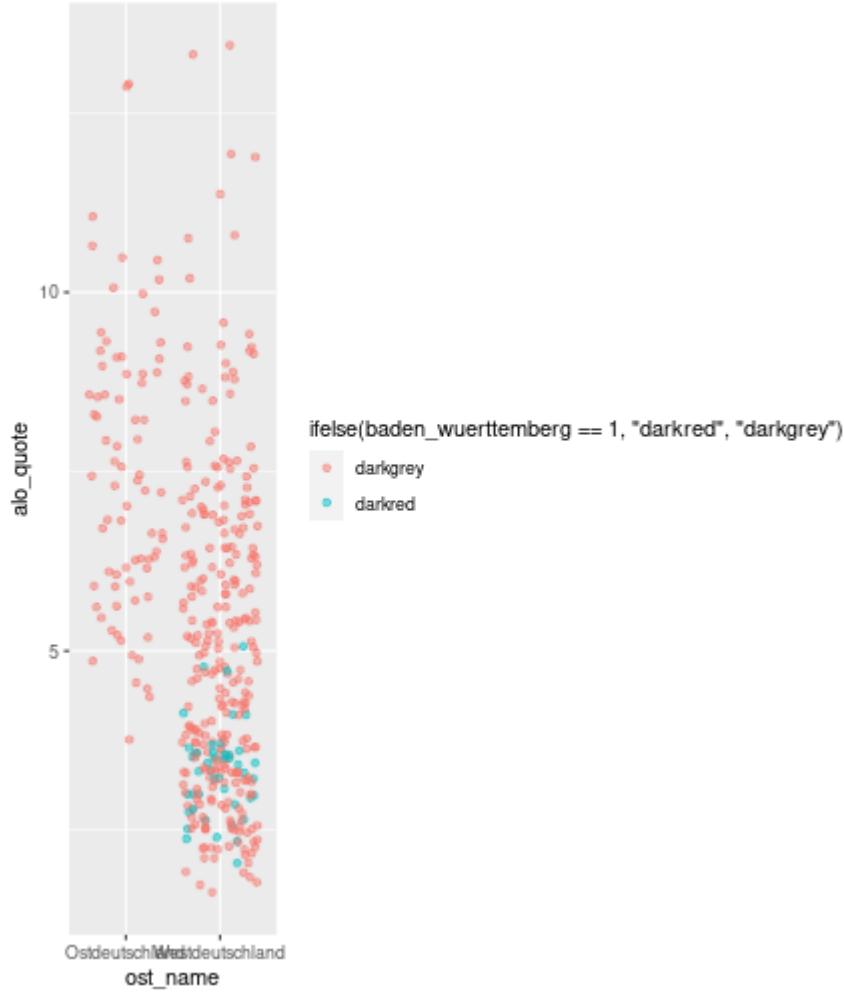
```
## # A tibble: 401 × 4
##   alo_quote landkreis_name     bundesland_name    ost_name
##       <dbl> <chr>           <chr>              <chr>
## 1      7.01 Flensburg        Schleswig-Holstein Westdeutschland
## 2      6.72 Kiel             Schleswig-Holstein Westdeutschland
## 3      7.04 Lübeck          Schleswig-Holstein Westdeutschland
## 4      6.82 Neumünster      Schleswig-Holstein Westdeutschland
## 5      7.11 Dithmarschen    Schleswig-Holstein Westdeutschland
## 6      7.43 Herzogtum Lauenburg Schleswig-Holstein Westdeutschland
## 7      5.90 Nordfriesland    Schleswig-Holstein Westdeutschland
## 8      5.97 Ostholstein      Schleswig-Holstein Westdeutschland
## 9      6.23 Pinneberg        Schleswig-Holstein Westdeutschland
## 10     6.92 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 × 5
##   alo_quote landkreis_name   bundesland_name    ost_name    baden_wuer
##       <dbl> <chr>           <chr>          <chr>        <fct>
## 1     7.01 Flensburg      Schleswig-Holstein Westdeutschland 0
## 2     6.72 Kiel           Schleswig-Holstein Westdeutschland 0
## 3     7.04 Lübeck         Schleswig-Holstein Westdeutschland 0
## 4     6.82 Neumünster    Schleswig-Holstein Westdeutschland 0
## 5     7.11 Dithmarschen   Schleswig-Holstein Westdeutschland 0
## 6     7.43 Herzogtum Lauenburg Schleswig-Holstein Westdeutschland 0
## 7     5.90 Nordfriesland   Schleswig-Holstein Westdeutschland 0
## 8     5.97 Ostholstein    Schleswig-Holstein Westdeutschland 0
## 9     6.23 Pinneberg      Schleswig-Holstein Westdeutschland 0
## 10    6.92 Plön           Schleswig-Holstein Westdeutschland 0
## # ... with 391 more rows, and abbreviated variable name `baden_wuerttemberg`
```

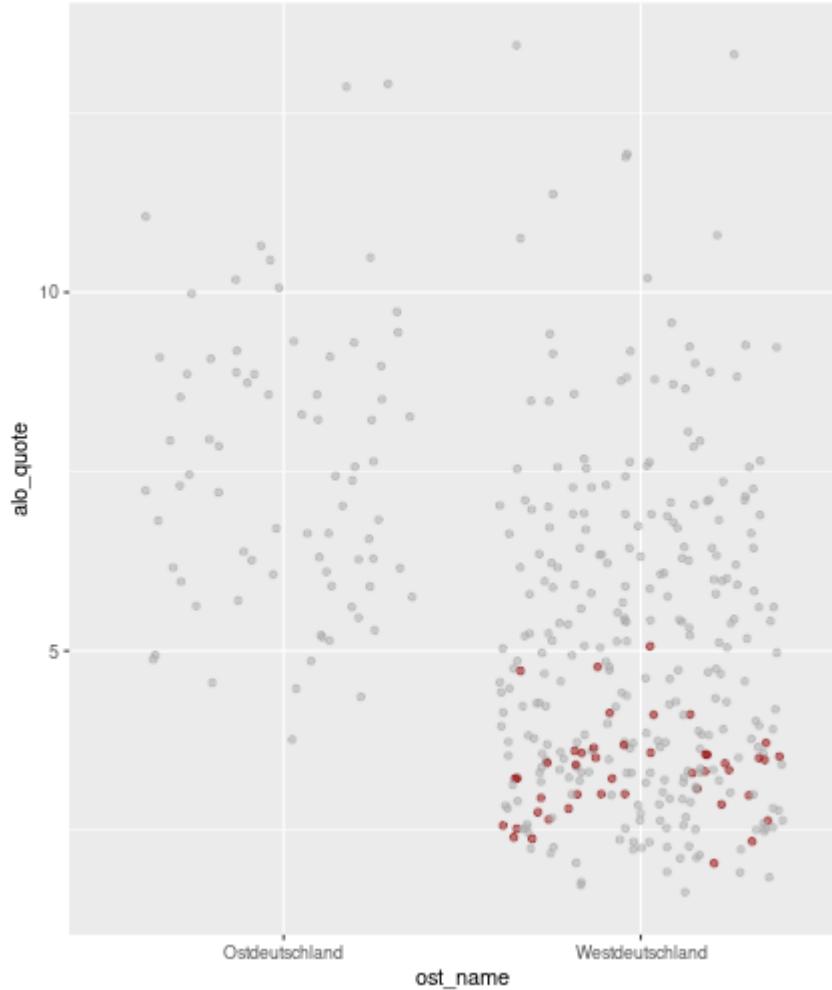
```
gesamtdataen %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
  ggplot(aes(x = ost_name, y=alo_quote))
```



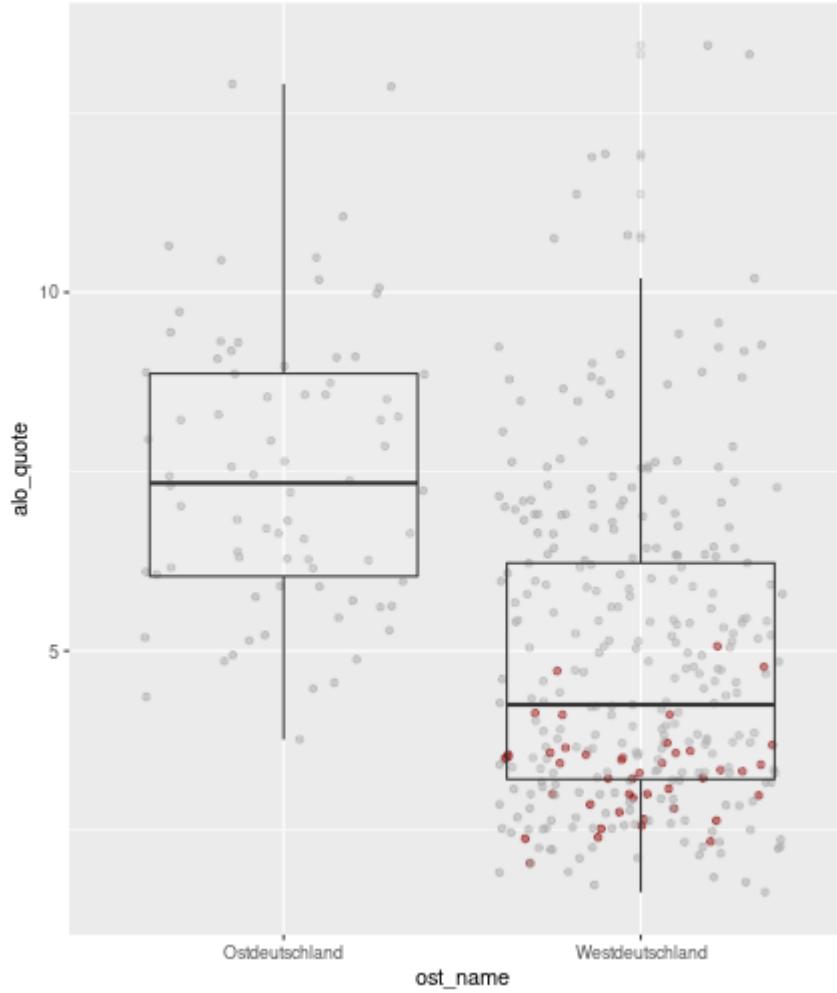
```
gesamtdataen %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
```



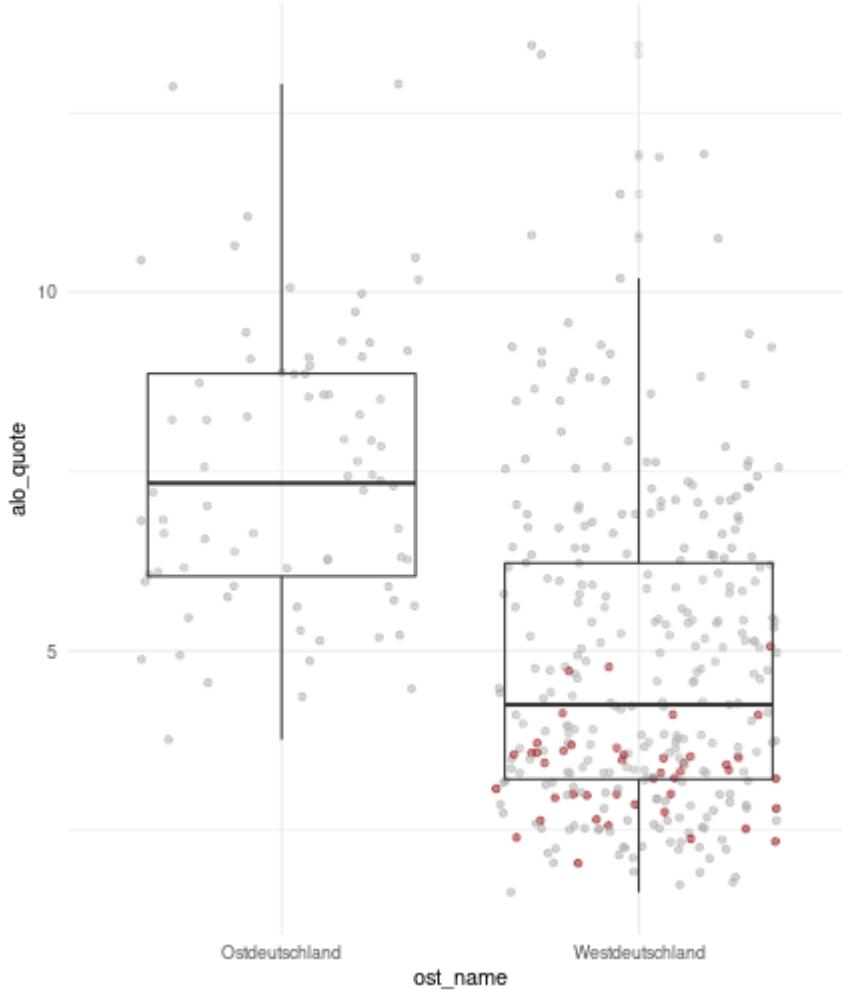
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity()
```



```
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()
```



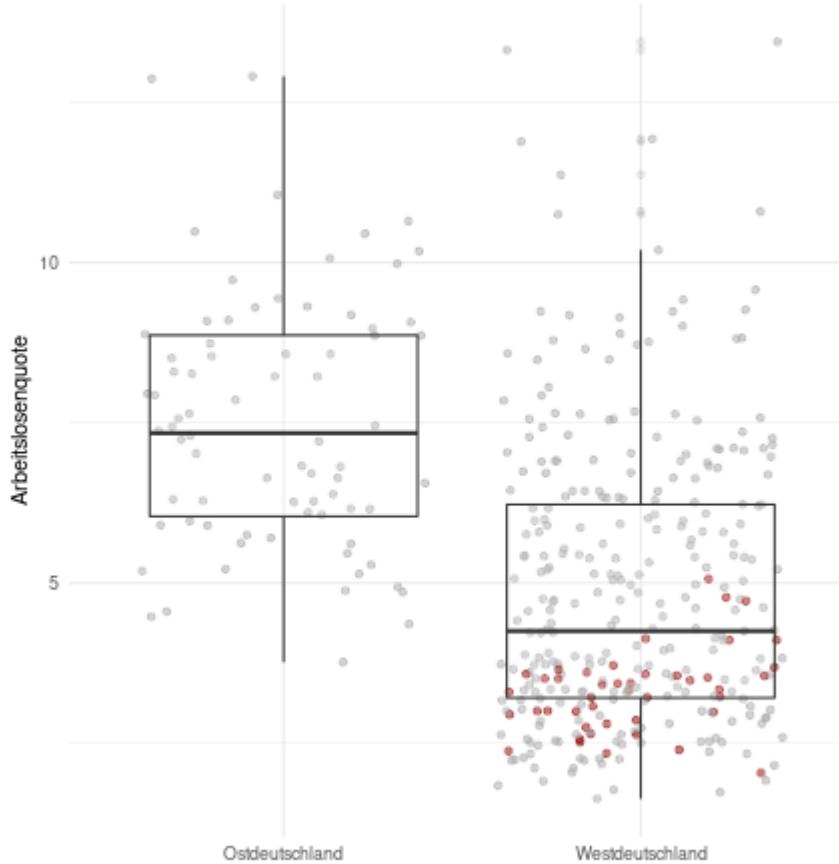
```

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



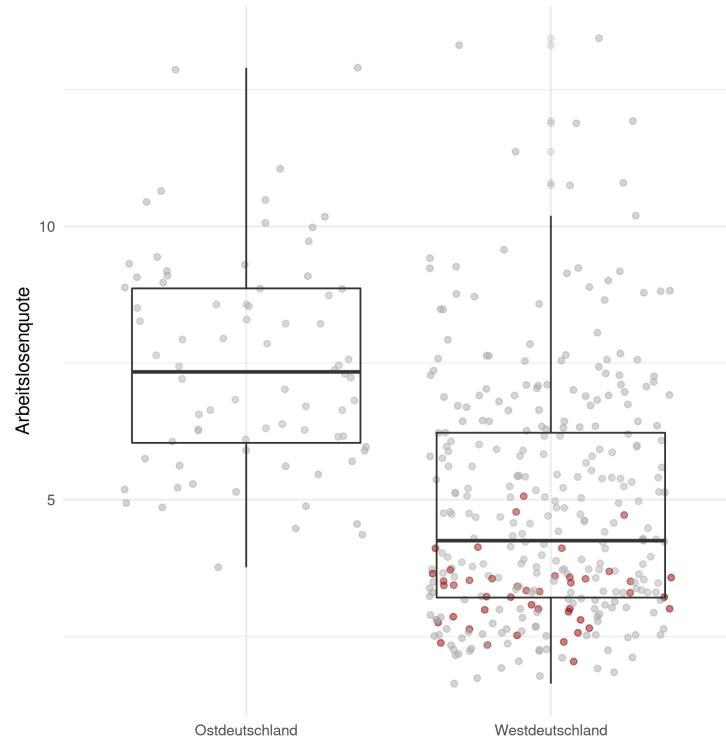
Quelle: Daten der Agentur für Arbeit aus dem Jahr 2017

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 Baden-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: 11,172 × 8

	Regionalschluessel	Jahr	bip	einwoh... ¹	bip_p... ²	bunde... ³	landk... ⁴	ost_r
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<chr>	<chr>
## 1	08111	1992	32946884000	593628	55501.	Baden-	Stuttg..	Westc
## 2	08115	1992	12090930000	343190	35231.	Baden-	Böblin..	Westc
## 3	08116	1992	12275605000	487370	25187.	Baden-	Esslin..	Westc
## 4	08117	1992	5062037000	248688	20355.	Baden-	Göppin..	Westc
## 5	08118	1992	11714160000	475248	24649.	Baden-	Ludwig..	Westc
## 6	08119	1992	8500405000	389670	21814.	Baden-	Rems-M..	Westc
## 7	08121	1992	4219259000	118566	35586.	Baden-	Heilbr..	Westc
## 8	08125	1992	6073525000	283163	21449.	Baden-	Heilbr..	Westc
## 9	08126	1992	2273334000	96072	23663.	Baden-	Hohenl..	Westc
## 10	08127	1992	3432175000	169617	20235.	Baden-	Schwäb..	Westc

... with 11,162 more rows, and abbreviated variable names ¹einwohner,
²bip_pro_kopf, ³bundesland_name, ⁴landkreis_name, ⁵ost_name

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

```
## # A tibble: 8,379 × 8
##   Regionalschluessel Jahr      bip einwoh...¹ bip_p...² bundes...³ landk...⁴ ost_r...
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>    <chr>    <chr>    <chr>
## 1 08111            2000 35273886000 571528 61719. Baden-Württemberg Stuttgart-Westen
## 2 08115            2000 13867882000 359476 38578. Baden-Württemberg Böblingen-Westen
## 3 08116            2000 14404617000 492914 29223. Baden-Württemberg Esslingen-Westen
## 4 08117            2000 6000420000 253970 23626. Baden-Württemberg Göppingen-Westen
## 5 08118            2000 14657540000 492014 29791. Baden-Württemberg Ludwigshafen-Westen
## 6 08119            2000 10367512000 403830 25673. Baden-Württemberg Rems-Murr-Kreis-Westen
## 7 08121            2000 5273634000 115590 45624. Baden-Württemberg Heilbronn-Westen
## 8 08125            2000 8453750000 316406 26718. Baden-Württemberg Heilbronn-Westen
## 9 08126            2000 3083030000 106494 28950. Baden-Württemberg Hohenlohe-Westen
## 10 08127           2000 4503834000 184222 24448. Baden-Württemberg Schwäbisch Hall-Westen
## # ... with 8,369 more rows, and abbreviated variable names `¹einwohner,
## #       `²bip_pro_kopf, `³bundesland_name, `⁴landkreis_name, `⁵ost_name
```

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

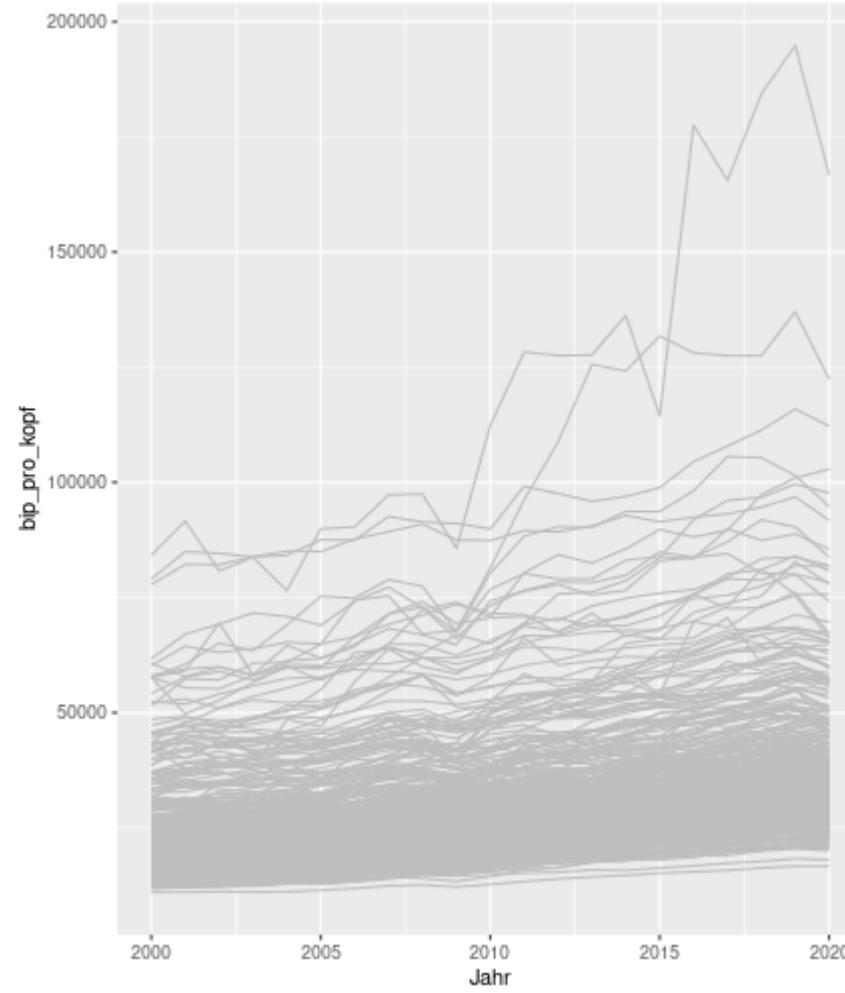
```
## # A tibble: 8,379 × 8
## # Groups:   ost_name, Jahr [42]
##   Regionalschluessel Jahr      bip einwoh...¹ bip_p...² bundes...³ landk...⁴ ost_r
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 08111             2000 35273886000  571528  61719. Baden-Württemberg Stuttgart-West...
## 2 08115             2000 13867882000  359476  38578. Baden-Württemberg Böblingen-West...
## 3 08116             2000 14404617000  492914  29223. Baden-Württemberg Esslingen-West...
## 4 08117             2000 60004200000  253970  23626. Baden-Württemberg Göppingen-West...
## 5 08118             2000 14657540000  492014  29791. Baden-Württemberg Ludwigshafen-West...
## 6 08119             2000 10367512000  403830  25673. Baden-Württemberg Rems-Murr-Kreis-West...
## 7 08121             2000 5273634000  115590  45624. Baden-Württemberg Heilbronn-West...
## 8 08125             2000 8453750000  316406  26718. Baden-Württemberg Heilbronn-West...
## 9 08126             2000 3083030000  106494  28950. Baden-Württemberg Hohenlohe-West...
## 10 08127            2000 4503834000  184222  24448. Baden-Württemberg Schwäbisch Hall-West...
## # ... with 8,369 more rows, and abbreviated variable names `¹einwohner,
## #       `²bip_pro_kopf, `³bundesland_name, `⁴landkreis_name, `⁵ost_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_
```

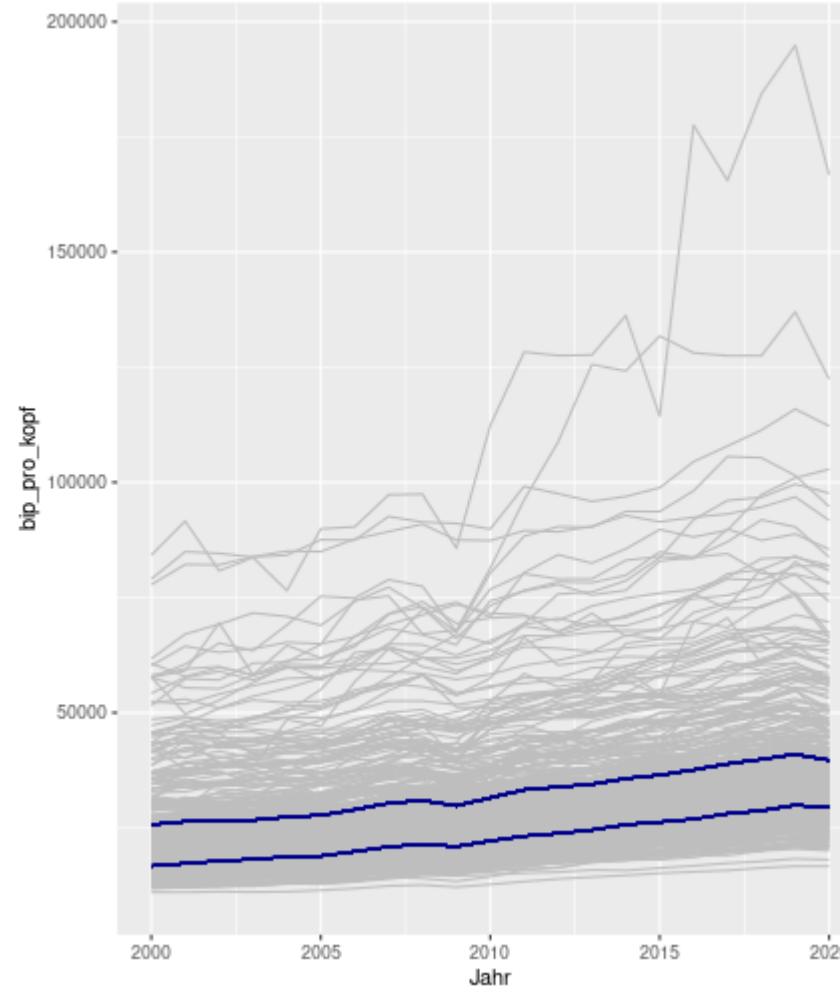
```
## # A tibble: 8,379 × 10
## # Groups:   ost_name, Jahr [42]
##   Regiona...¹ Jahr     bip_einwo...² bip_p...³ bunde...⁴ landk...⁵ ost_n...⁶ durch...⁷ ...
##   <chr>      <dbl>    <dbl>    <dbl>    <chr>    <chr>    <chr>    <dbl>    <dbl>
## 1 08111    2000 3.53e10  571528  61719. Baden-... Stuttg... Westde... 25714.
## 2 08115    2000 1.39e10  359476  38578. Baden-... Böblin... Westde... 25714.
## 3 08116    2000 1.44e10  492914  29223. Baden-... Esslin... Westde... 25714.
## 4 08117    2000 6.00e 9  253970  23626. Baden-... Göppin... Westde... 25714.
## 5 08118    2000 1.47e10  492014  29791. Baden-... Ludwig... Westde... 25714.
## 6 08119    2000 1.04e10  403830  25673. Baden-... Rems-M... Westde... 25714.
## 7 08121    2000 5.27e 9  115590  45624. Baden-... Heilbr... Westde... 25714.
## 8 08125    2000 8.45e 9  316406  26718. Baden-... Heilbr... Westde... 25714.
## 9 08126    2000 3.08e 9  106494  28950. Baden-... Hohenl... Westde... 25714.
## 10 08127   2000 4.50e 9  184222  24448. Baden-... Schwäb... Westde... 25714.
## # ... with 8,369 more rows, and abbreviated variable names ¹Regionalschluessel,
## #       ²einwohner, ³bip_pro_kopf, ⁴bundesland_name, ⁵landkreis_name, ⁶ost_name,
## #       ⁷durchschnitt
```

```
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()
```

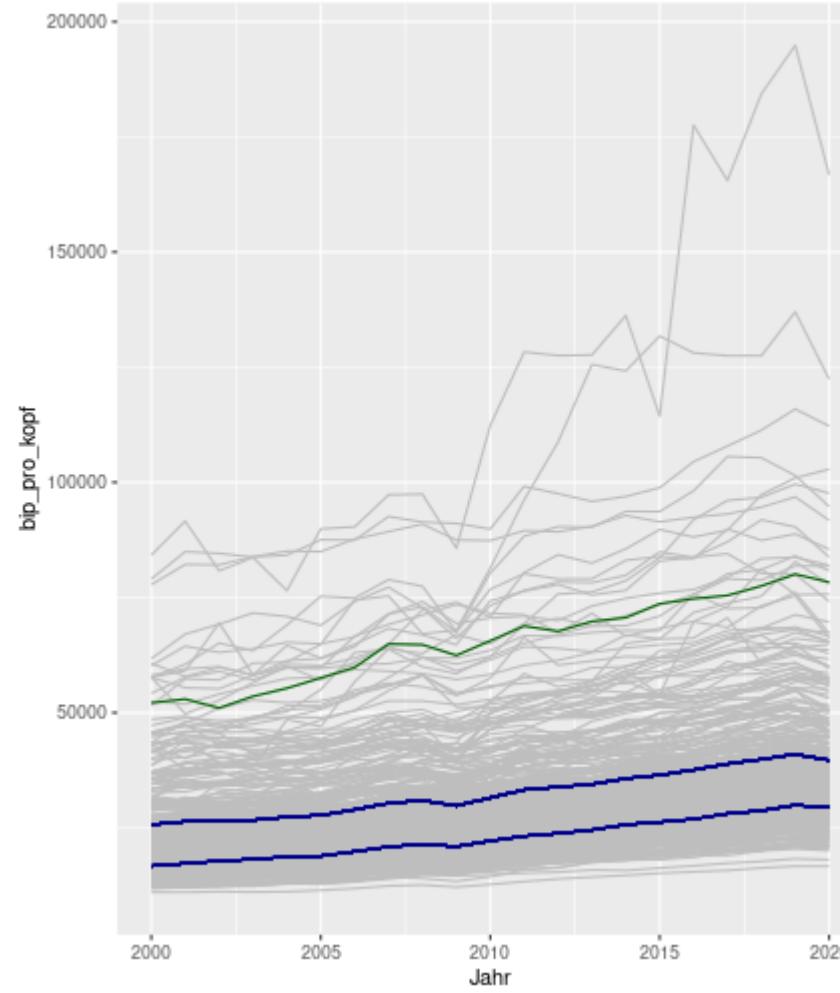
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0),
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
```



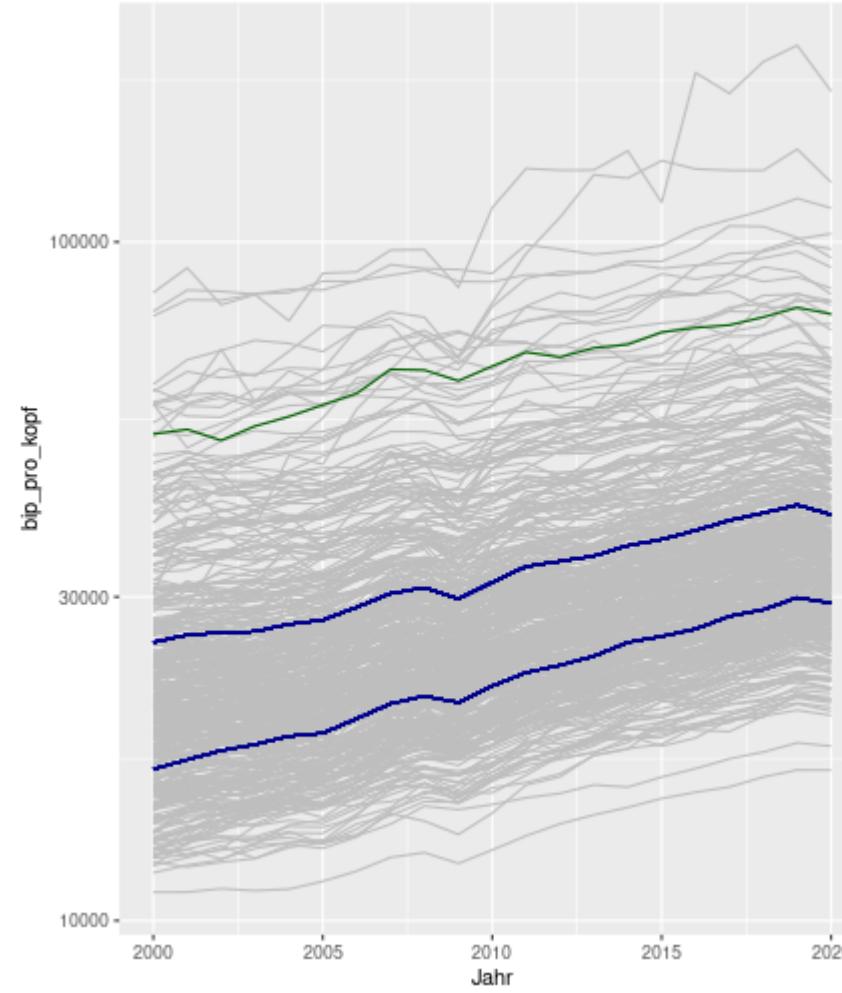
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0))
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x =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



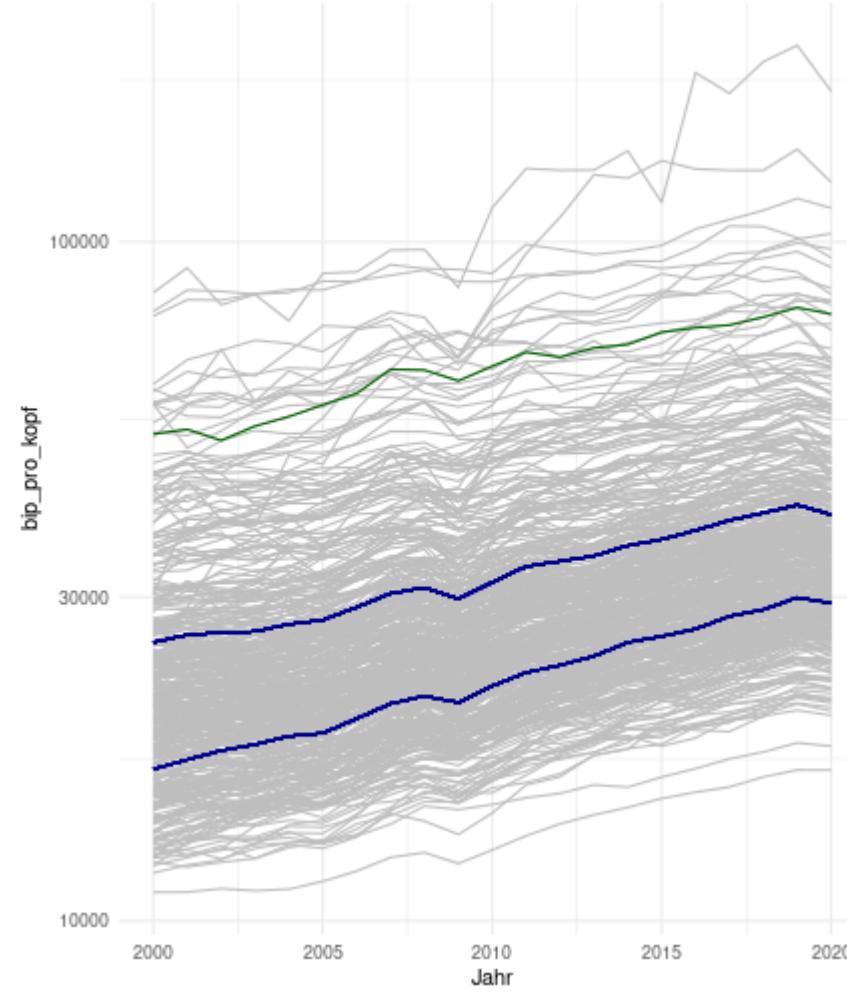
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr,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



```
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")
```



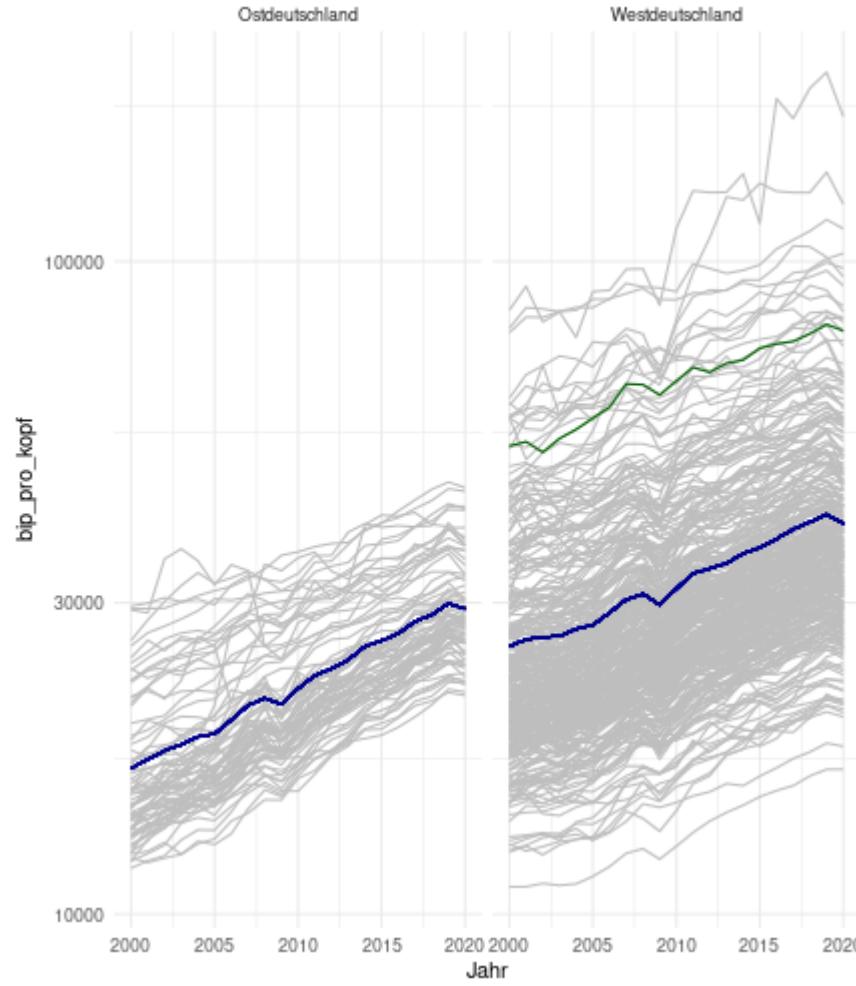
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr,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

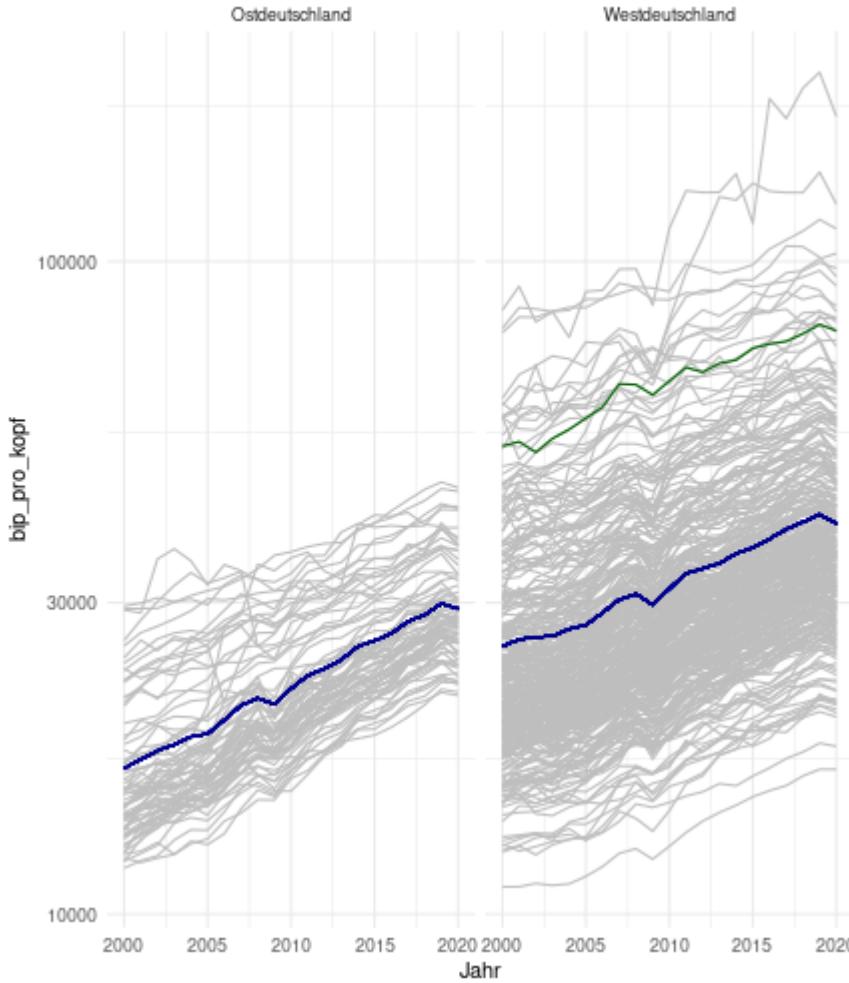
```



```

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")

```

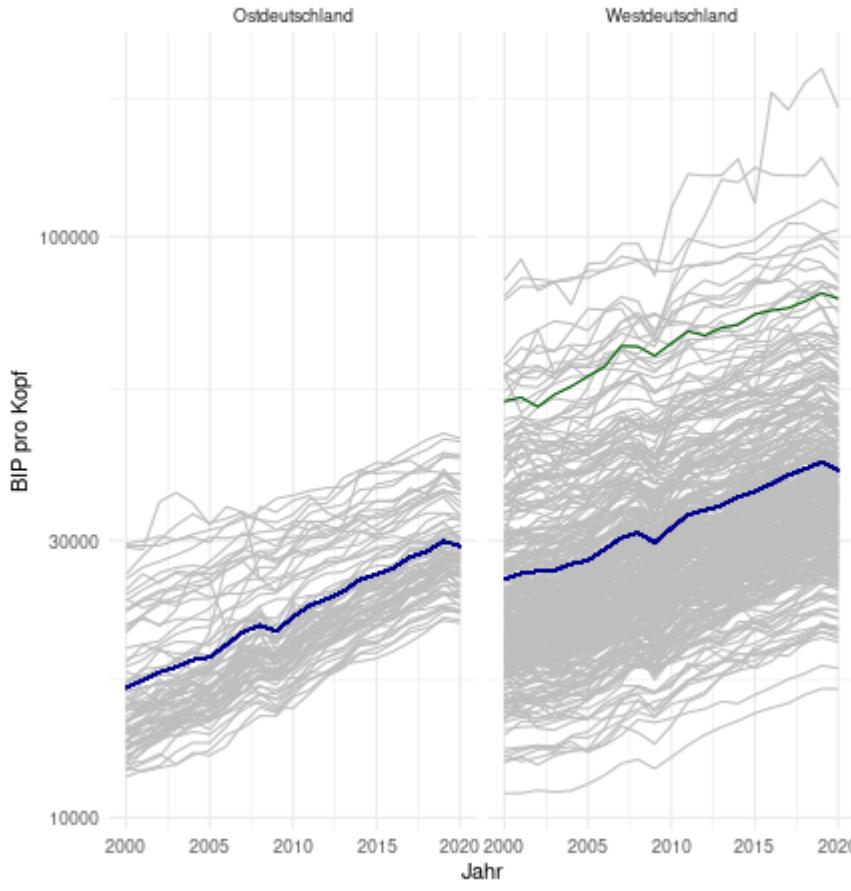


```

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 Landkreisen 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 den ostdeutschen Landkreisen zu, doch erreicht es mit durchschnittlich 28091€ 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 Landkreisen 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 den ostdeutschen Landkreisen zu, doch erreicht es mit durchschnittlich 28091€ 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:

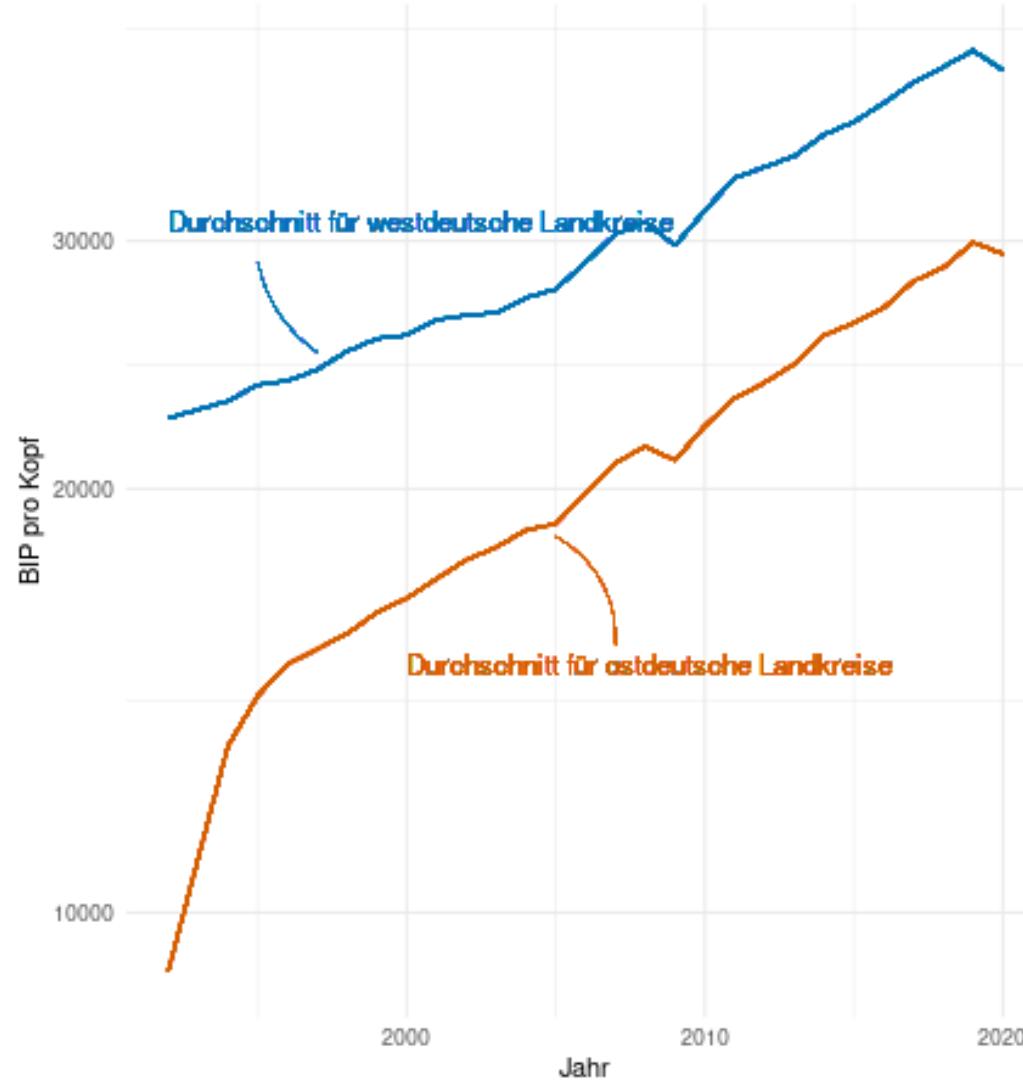
- + Eine Wachstumsprozess im BIP pro Kopf findet in allen Landkreisen statt, jedoch gibt es für die ostdeutschen Landkreise, welche deutlich niedriger gestartet sind, keinen erkennbaren Anpassungsprozess in Form eines schnelleren Wachstums
- + Wir sehen auch keinen Anpassungsprozess der Landkreise in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

Bruttoinlandsprodukt pro Kopf

Daten ab 1992 vorhanden, d.h. wir können auch weiter zurück gehen:

- ✚ Allerdings: Keine Daten zu *allen* Landkreisen, daher Vorsicht!
- ✚ Hier sehen wir einen Anpassungsprozess in den 1990er Jahren
- ✚ Anpassung verlangsamt sich, ab 2010 praktisch parallel

Ein Vergleich des BIP pro Kopf von ost- und westdeutschen Landkreisen
Zeitreihe ab 1992 bis 2017



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

Wachstum des BIP pro Kopf

Paneldaten beim BIP pro Kopf vorhanden, d.h. wir können:

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

berechnen und visualisieren.

Wachstum des BIP pro Kopf

Paneldaten beim BIP pro Kopf vorhanden, d.h. wir können:

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

berechnen und visualisieren.

| Können wir einen Anpassungsprozess über die Wachstumsraten des BIP pro Kopf feststellen?

```
bip_zeitreihe_namen
```

```
## # A tibble: 11,172 × 8
##   Regionalschluessel Jahr      bip einwoh...¹ bip_p...² bunde...³ landk...⁴ ost_r
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 08111            1992 32946884000  593628 55501. Baden-... Stuttg... Westc
## 2 08115            1992 12090930000  343190 35231. Baden-... Böblin... Westc
## 3 08116            1992 12275605000  487370 25187. Baden-... Esslin... Westc
## 4 08117            1992 5062037000  248688 20355. Baden-... Göppin... Westc
## 5 08118            1992 11714160000  475248 24649. Baden-... Ludwigs... Westc
## 6 08119            1992 8500405000  389670 21814. Baden-... Rems-M... Westc
## 7 08121            1992 4219259000  118566 35586. Baden-... Heilbr... Westc
## 8 08125            1992 6073525000  283163 21449. Baden-... Heilbr... Westc
## 9 08126            1992 2273334000  96072 23663. Baden-... Hohenlo... Westc
## 10 08127           1992 3432175000  169617 20235. Baden-... Schwäb... Westc
## # ... with 11,162 more rows, and abbreviated variable names `einwohner,
## #       `bip_pro_kopf`, `bundesland_name`, `landkreis_name`, `ost_name`
```

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

```
## # A tibble: 11,172 × 8
## # Groups:   Regionalschluessel [399]
##   Regionalschluessel Jahr      bip einwoh...¹ bip_p...² bundes...³ landk...⁴ ost_r...
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl> <chr>    <chr>    <chr>
## 1 08111            1992 32946884000  593628 55501. Baden-Württemberg Stuttgart-West...
## 2 08115            1992 12090930000  343190 35231. Baden-Württemberg Böblingen-West...
## 3 08116            1992 12275605000  487370 25187. Baden-Württemberg Esslingen-West...
## 4 08117            1992 5062037000  248688 20355. Baden-Württemberg Göppingen-West...
## 5 08118            1992 11714160000  475248 24649. Baden-Württemberg Ludwigshafen-West...
## 6 08119            1992 8500405000  389670 21814. Baden-Württemberg Rems-Murr-Kreis-West...
## 7 08121            1992 4219259000  118566 35586. Baden-Württemberg Heilbronn-West...
## 8 08125            1992 6073525000  283163 21449. Baden-Württemberg Heilbronn-West...
## 9 08126            1992 2273334000  96072 23663. Baden-Württemberg Hohenlohe-West...
## 10 08127           1992 3432175000  169617 20235. Baden-Württemberg Schwäbisch Hall-West...
## # ... with 11,162 more rows, and abbreviated variable names `¹einwohner,
## #       `²bip_pro_kopf, `³bundesland_name, `⁴landkreis_name, `⁵ost_name
```

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

## # A tibble: 11,172 × 8
## # Groups:   Regionalschluessel [399]
##   Regionalschluessel Jahr     bip_einwohner bip_p...¹ bundes...² landk...³ ost_r...
##   <chr>           <dbl>        <dbl>        <dbl> <dbl> <chr>    <chr>    <chr>
## 1 01001            1992 25555393000 86642 29494. Schles... Flensb... Westc...
## 2 01001            1994 2624290000 86287 30414. Schles... Flensb... Westc...
## 3 01001            1995 2628469000 85506 30740. Schles... Flensb... Westc...
## 4 01001            1996 2578880000 84499 30520. Schles... Flensb... Westc...
## 5 01001            1997 2725086000 83344 32697. Schles... Flensb... Westc...
## 6 01001            1998 2872187000 82112 34979. Schles... Flensb... Westc...
## 7 01001            1999 2712126000 81276 33369. Schles... Flensb... Westc...
## 8 01001            2000 2487282000 80758 30799. Schles... Flensb... Westc...
## 9 01001            2001 2465393000 80489 30630. Schles... Flensb... Westc...
## 10 01001           2002 2635779000 80414 32778. Schles... Flensb... Westc...
## # ... with 11,162 more rows, and abbreviated variable names `¹`bip_pro_kopf,
## #       `²`bundesland_name, `³`landkreis_name, `⁴`ost_name
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
```

A tibble: 11,172 × 9
Groups: Regionalschluessel [399]
Regionalschlu...¹ Jahr bip_einwo...² bip_p...³ bunde...⁴ landk...⁵ ost_n...⁶ bip_r...⁷
<chr> <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl>
1 01001 1992 2.56e9 86642 29494. Schles... Flensb... Westde... NA
2 01001 1994 2.62e9 86287 30414. Schles... Flensb... Westde... 3.0
3 01001 1995 2.63e9 85506 30740. Schles... Flensb... Westde... 1.0
4 01001 1996 2.58e9 84499 30520. Schles... Flensb... Westde... -0.7
5 01001 1997 2.73e9 83344 32697. Schles... Flensb... Westde... 6.6
6 01001 1998 2.87e9 82112 34979. Schles... Flensb... Westde... 6.5
7 01001 1999 2.71e9 81276 33369. Schles... Flensb... Westde... -4.8
8 01001 2000 2.49e9 80758 30799. Schles... Flensb... Westde... -8.3
9 01001 2001 2.47e9 80489 30630. Schles... Flensb... Westde... -0.5
10 01001 2002 2.64e9 80414 32778. Schles... Flensb... Westde... 6.5
... with 11,162 more rows, and abbreviated variable names ¹Regionalschluessel,
²einwohner, ³bip_pro_kopf, ⁴bundesland_name, ⁵landkreis_name, ⁶ost_name,
⁷bip_pro_kopf_wachstum

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup()

## # A tibble: 11,172 × 9
##   Regionalschluessel Jahr   bip_einwo...¹ bip_p...² bunde...⁴ landk...⁵ ost_n...⁶ bip_p...
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl>    <chr>    <chr>    <chr>    <dbl>
## 1 01001            1992  2.56e9   86642  29494. Schles... Flensb... Westde... NA
## 2 01001            1994  2.62e9   86287  30414. Schles... Flensb... Westde... 3.0
## 3 01001            1995  2.63e9   85506  30740. Schles... Flensb... Westde... 1.0
## 4 01001            1996  2.58e9   84499  30520. Schles... Flensb... Westde... -0.7
## 5 01001            1997  2.73e9   83344  32697. Schles... Flensb... Westde... 6.6
## 6 01001            1998  2.87e9   82112  34979. Schles... Flensb... Westde... 6.5
## 7 01001            1999  2.71e9   81276  33369. Schles... Flensb... Westde... -4.8
## 8 01001            2000  2.49e9   80758  30799. Schles... Flensb... Westde... -8.3
## 9 01001            2001  2.47e9   80489  30630. Schles... Flensb... Westde... -0.5
## 10 01001           2002  2.64e9   80414  32778. Schles... Flensb... Westde... 6.5
## # ... with 11,162 more rows, and abbreviated variable names `¹Regionalschluessel,
## #       `²einwohner, `³bip_pro_kopf, `⁴bundesland_name, `⁵landkreis_name, `⁶ost_name,
## #       `⁷bip_pro_kopf_wachstum
```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr)
```

```
## # A tibble: 11,172 × 9
## # Groups: ost_name, Jahr [56]
##   Regionalschlu...¹ Jahr   bip_einwo...² bip_p...³ bundes...⁴ landk...⁵ ost_n...⁶ bip_r...
##   <chr>           <dbl>   <dbl>   <dbl>   <dbl>   <chr>   <chr>   <chr>   <dbl>
## 1 01001            1992 2.56e9  86642  29494. Schles... Flensb... Westde... NA
## 2 01001            1994 2.62e9  86287  30414. Schles... Flensb... Westde... 3.0
## 3 01001            1995 2.63e9  85506  30740. Schles... Flensb... Westde... 1.0
## 4 01001            1996 2.58e9  84499  30520. Schles... Flensb... Westde... -0.7
## 5 01001            1997 2.73e9  83344  32697. Schles... Flensb... Westde... 6.6
## 6 01001            1998 2.87e9  82112  34979. Schles... Flensb... Westde... 6.5
## 7 01001            1999 2.71e9  81276  33369. Schles... Flensb... Westde... -4.8
## 8 01001            2000 2.49e9  80758  30799. Schles... Flensb... Westde... -8.3
## 9 01001            2001 2.47e9  80489  30630. Schles... Flensb... Westde... -0.5
## 10 01001           2002 2.64e9  80414  32778. Schles... Flensb... Westde... 6.5
## # ... with 11,162 more rows, and abbreviated variable names `¹Regionalschluessel,
## #       `²einwohner, `³bip_pro_kopf, `⁴bundesland_name, `⁵landkreis_name, `⁶ost_name,
## #       `⁷bip_pro_kopf_wachstum
```

```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
## # A tibble: 11,172 × 10
## # Groups:   ost_name, Jahr [56]
##   Region...¹ Jahr     bip_einwo...² bip_p...³ bundes...⁴ landk...⁵ ost_n...⁶ bip_p...⁷ durch...
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>  <chr>    <chr>    <dbl>    <dbl>    <dbl>
## 1 01001     1992 2.56e9  86642  29494. Schles... Flensb... Westde... NA     NaN
## 2 01001     1994 2.62e9  86287  30414. Schles... Flensb... Westde... 3.02   2.8
## 3 01001     1995 2.63e9  85506  30740. Schles... Flensb... Westde... 1.06   2.4
## 4 01001     1996 2.58e9  84499  30520. Schles... Flensb... Westde... -0.723  0.9
## 5 01001     1997 2.73e9  83344  32697. Schles... Flensb... Westde... 6.66   1.7
## 6 01001     1998 2.87e9  82112  34979. Schles... Flensb... Westde... 6.52   2.8
## 7 01001     1999 2.71e9  81276  33369. Schles... Flensb... Westde... -4.82   1.9
## 8 01001     2000 2.49e9  80758  30799. Schles... Flensb... Westde... -8.34   2.4
## 9 01001     2001 2.47e9  80489  30630. Schles... Flensb... Westde... -0.552  2.1
## 10 01001    2002 2.64e9  80414  32778. Schles... Flensb... Westde... 6.55   0.7
## # ... with 11,162 more rows, and abbreviated variable names `¹Regionalschluessel,
## #       `²einwohner, `³bip_pro_kopf, `⁴bundesland_name, `⁵landkreis_name, `⁶ost_name,
## #       `⁷bip_pro_kopf_wachstum, `⁸durchschnitt

```

```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum
```

```
bip_zeitreihe_namen %>%  
  group_by(Regionalschluessel) %>%  
  arrange(Regionalschluessel, Jahr) %>%  
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf  
  ungroup() %>%  
  group_by(ost_name, Jahr) %>%  
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,  
  ungroup() -> bip_wachstum
```

```

## # A tibble: 11,172 × 10
##   Region...¹ Jahr     bip_einwo...² bip_p...³ bundes...⁴ landk...⁵ ost_n...⁶ bip_p...⁷ durch...
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <chr>    <chr>    <chr>    <dbl>    <dbl>
## 1 01001      1992  2.56e9  86642  29494. Schles... Flensb... Westde... NA     NaN
## 2 01001      1994  2.62e9  86287  30414. Schles... Flensb... Westde... 3.02   2.8
## 3 01001      1995  2.63e9  85506  30740. Schles... Flensb... Westde... 1.06   2.4
## 4 01001      1996  2.58e9  84499  30520. Schles... Flensb... Westde... -0.723  0.9
## 5 01001      1997  2.73e9  83344  32697. Schles... Flensb... Westde... 6.66   1.1
## 6 01001      1998  2.87e9  82112  34979. Schles... Flensb... Westde... 6.52   2.8
## 7 01001      1999  2.71e9  81276  33369. Schles... Flensb... Westde... -4.82   1.9
## 8 01001      2000  2.49e9  80758  30799. Schles... Flensb... Westde... -8.34   2.4
## 9 01001      2001  2.47e9  80489  30630. Schles... Flensb... Westde... -0.552  2.1
## 10 01001     2002  2.64e9  80414  32778. Schles... Flensb... Westde... 6.55   0.7
## # ... with 11,162 more rows, and abbreviated variable names ¹Regionalschluessel,
## #   ²einwohner, ³bip_pro_kopf, ⁴bundesland_name, ⁵landkreis_name, ⁶ost_name,
## #   ⁷bip_pro_kopf_wachstum, ⁸durchschnitt

```

```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

bip_wachstum %>%
  filter( Jahr >= 2000 )

```

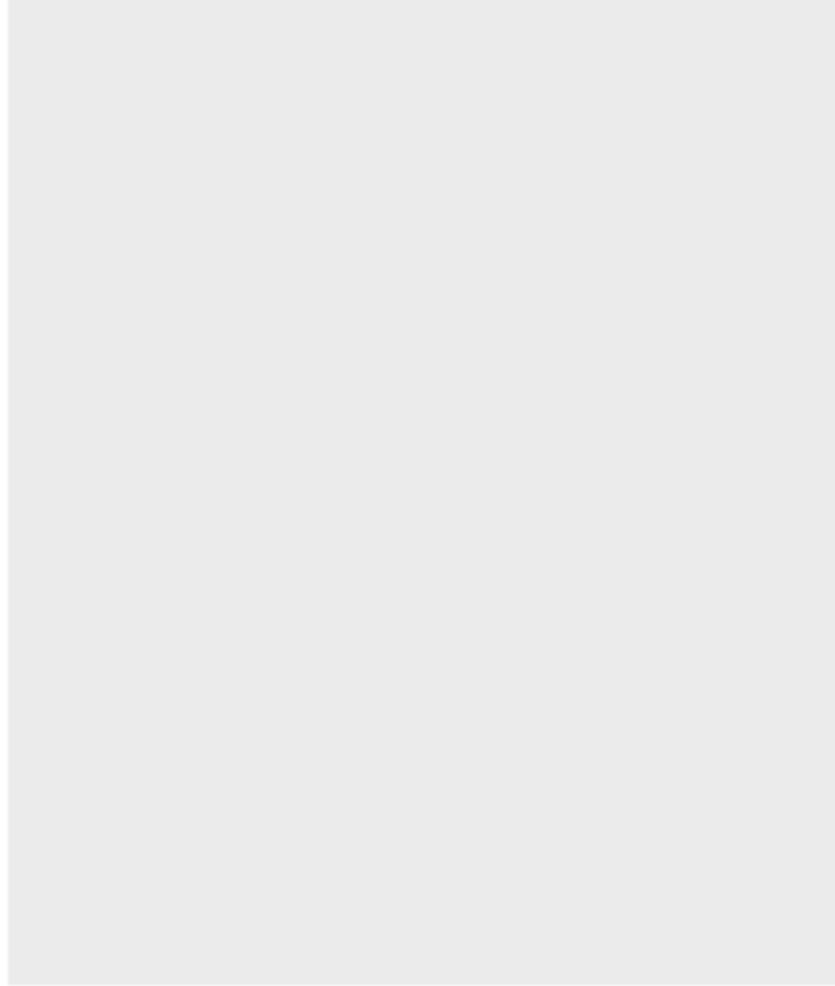
A tibble: 8,379 × 10

Region... ¹	Jahr	bip_einwo... ²	bip_p... ³	bunde... ⁴	landk... ⁵	ost_n... ⁶	bip_p... ⁷	durch...	
1 01001	2000	2.49e9	80758	30799.	Schles...	Flensb...	Westde...	-8.34	2.4
2 01001	2001	2.47e9	80489	30630.	Schles...	Flensb...	Westde...	-0.552	2.1
3 01001	2002	2.64e9	80414	32778.	Schles...	Flensb...	Westde...	6.55	0.7
4 01001	2003	2.71e9	80538	33590.	Schles...	Flensb...	Westde...	2.42	0.2
5 01001	2004	2.87e9	80783	35500.	Schles...	Flensb...	Westde...	5.38	2.3
6 01001	2005	2.90e9	80892	35860.	Schles...	Flensb...	Westde...	1.00	1.2
7 01001	2006	3.02e9	81052	37283.	Schles...	Flensb...	Westde...	3.82	4.1
8 01001	2007	3.05e9	81634	37377.	Schles...	Flensb...	Westde...	0.251	4.2
9 01001	2008	3.11e9	82403	37692.	Schles...	Flensb...	Westde...	0.835	2.2
10 01001	2009	2.99e9	82478	36300.	Schles...	Flensb...	Westde...	-3.83	-3.9

... with 8,369 more rows, and abbreviated variable names ¹Regionalschluessel,
²einwohner, ³bip_pro_kopf, ⁴bundesland_name, ⁵landkreis_name, ⁶ost_name,
⁷bip_pro_kopf_wachstum, ⁸durchschnitt

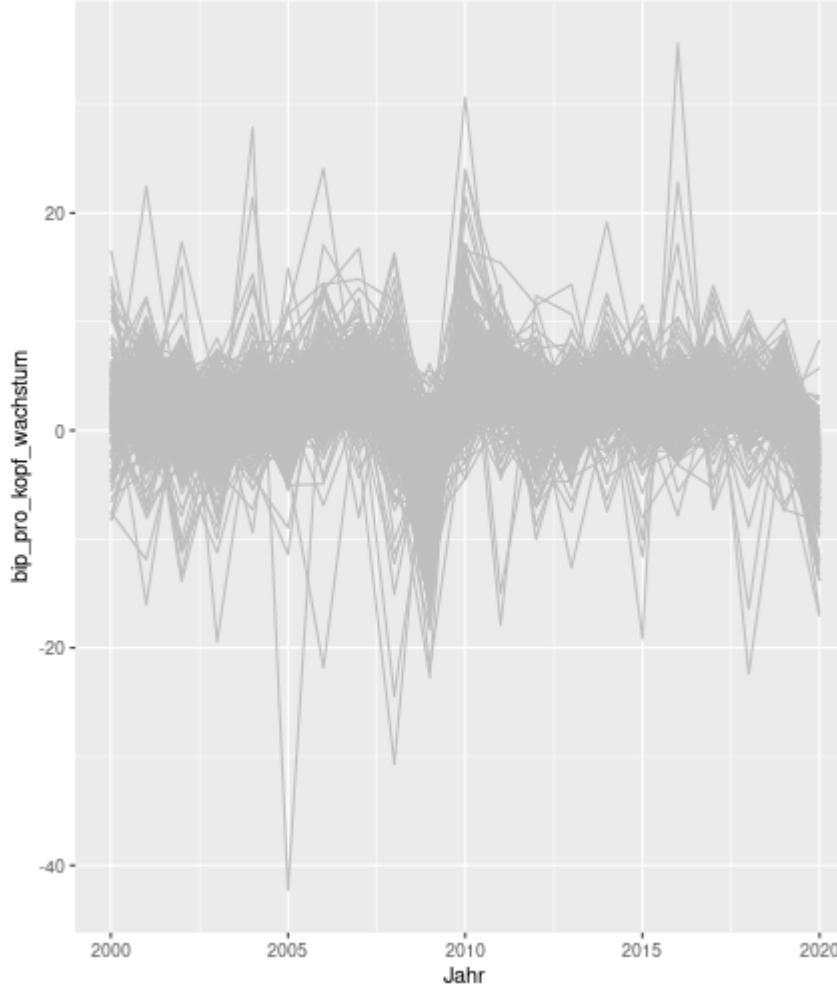
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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



```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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

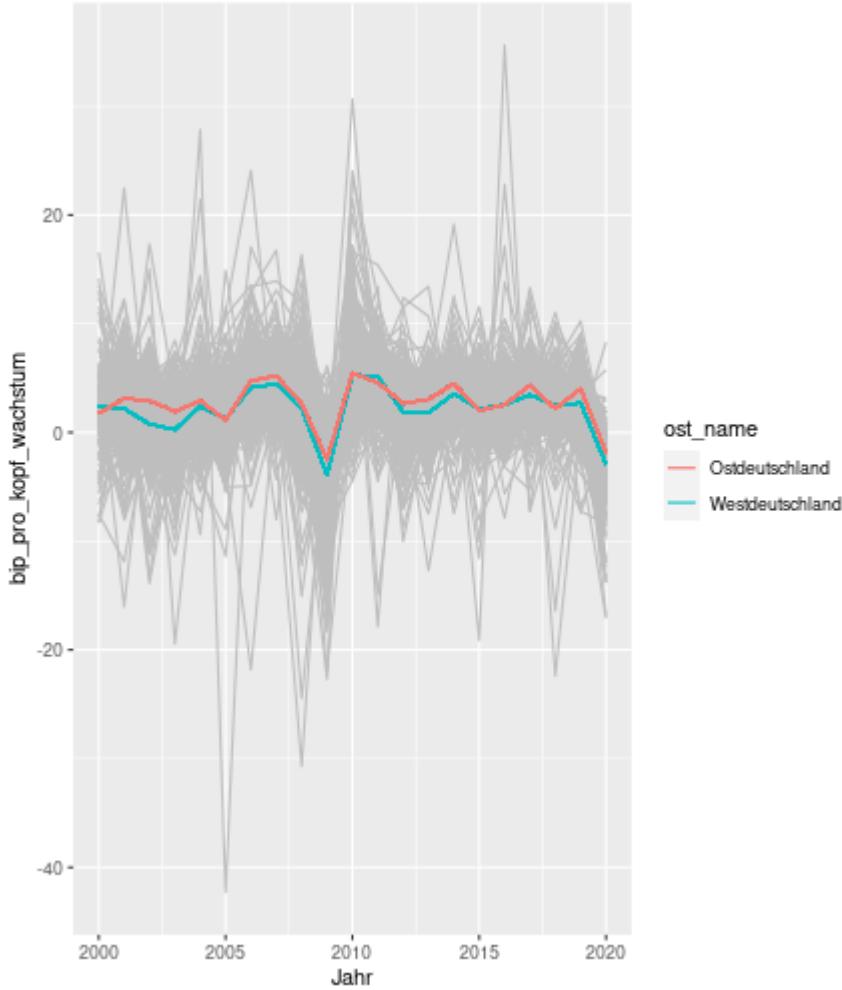


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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

```

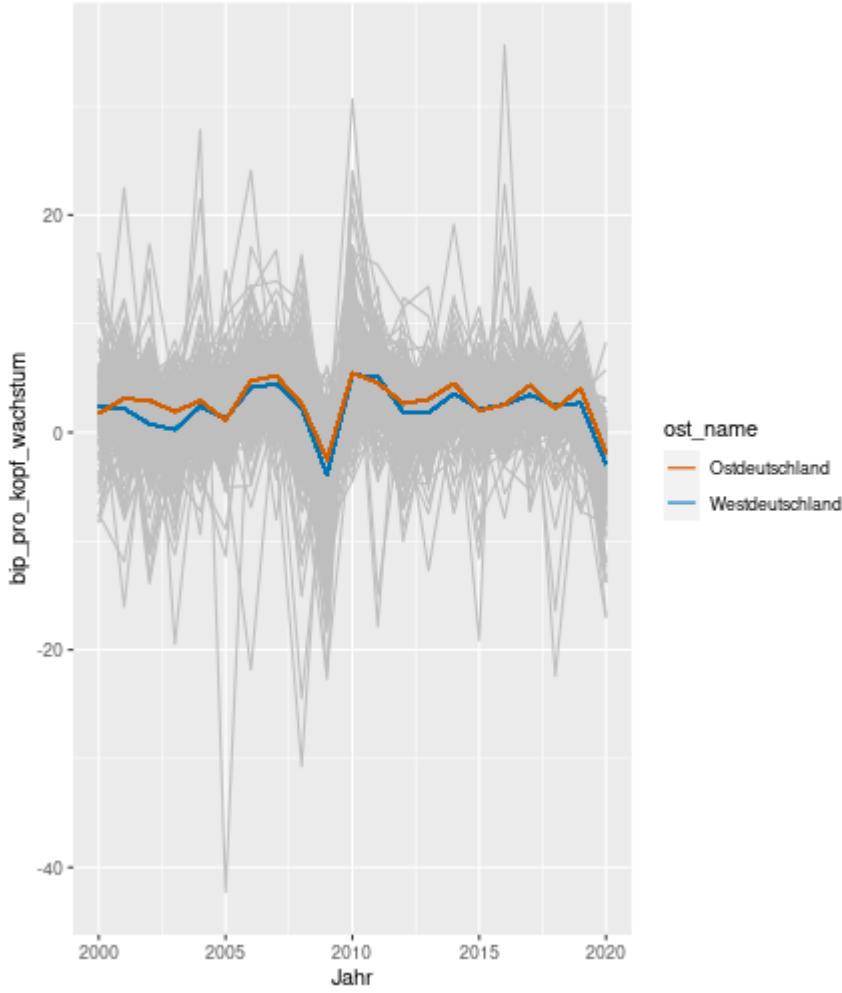


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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

```

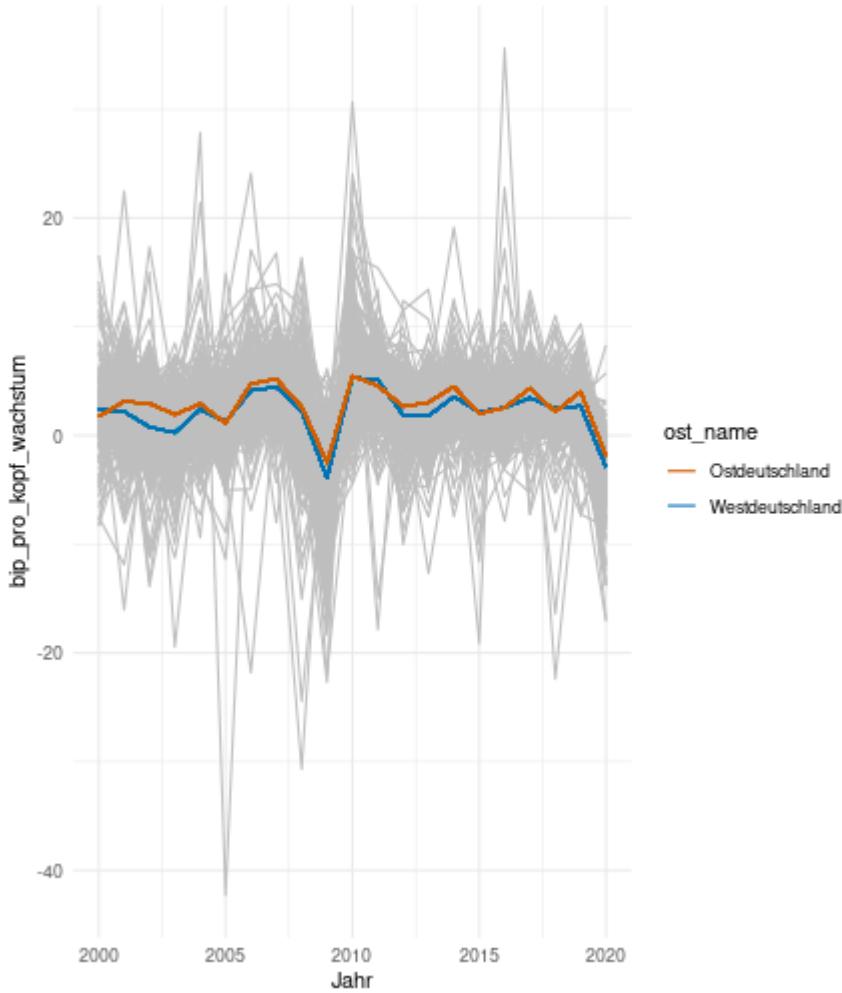


```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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

```



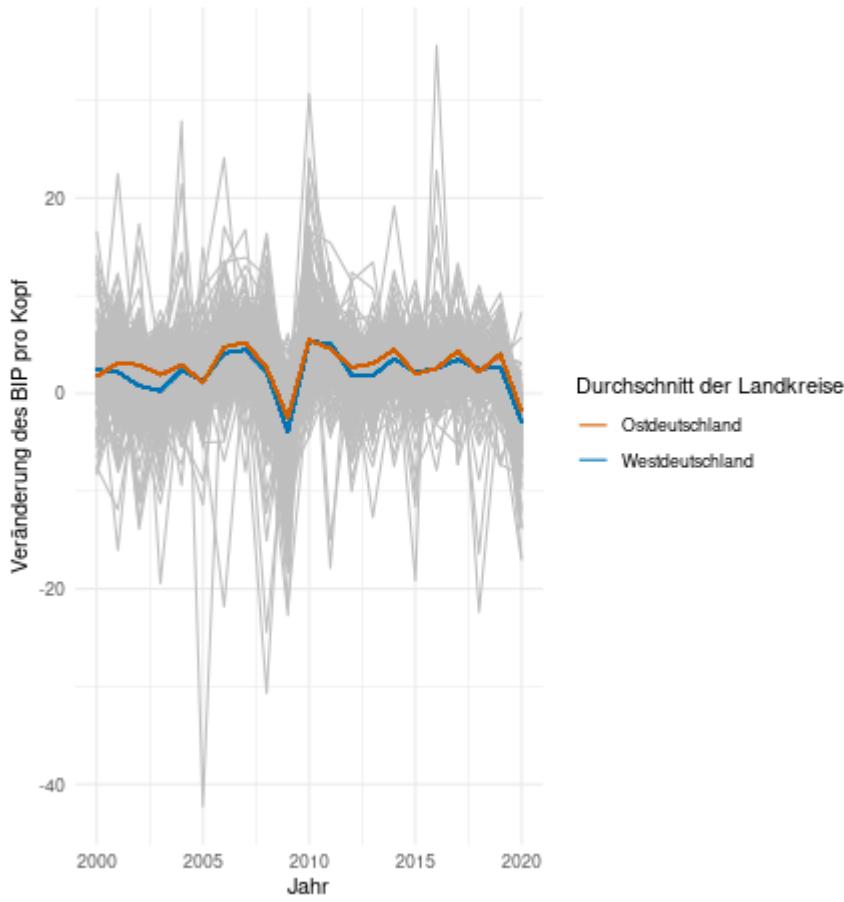
```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf_wachstum,
  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 pro Kopf v
  caption = "Quelle: Daten der Statistischen Äm
  x = "Jahr",
  y = "Veränderung des BIP pro Kopf")

```

Die Wachstumsrate des BIP pro Kopf von ost- und westdeutschen Landkreisen



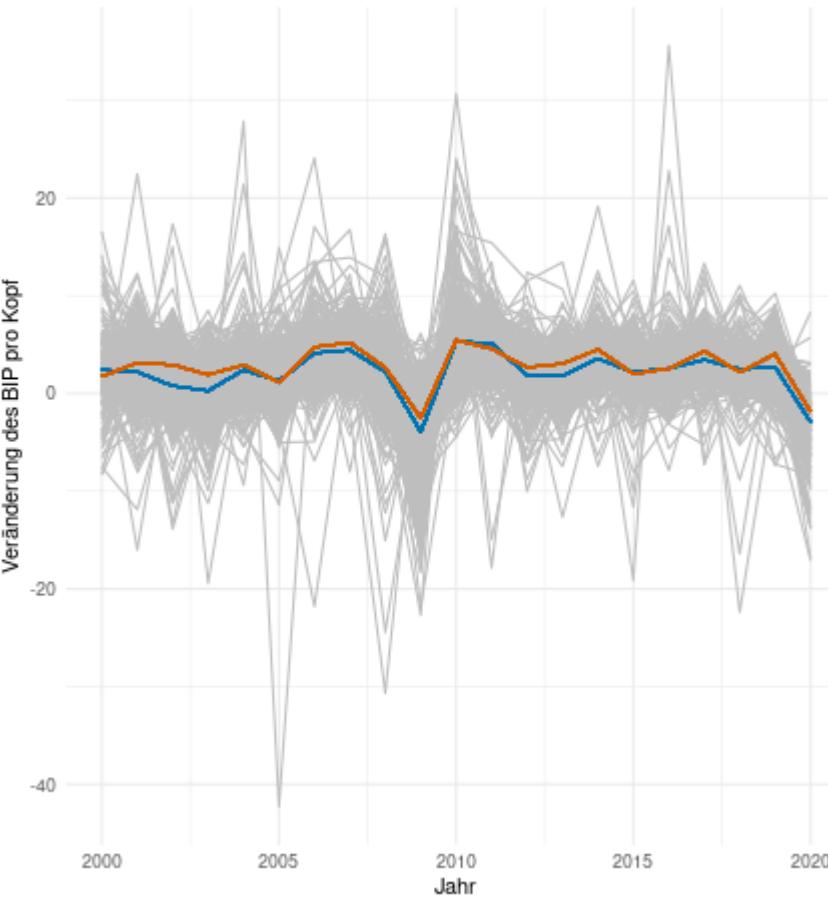
```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf_wachstum,
  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 pro Kopf v
  caption = "Quelle: Daten der Statistischen Äm
  x = "Jahr",
  y = "Veränderung des BIP pro Kopf") +
  theme(legend.position = "none")

```

Die Wachstumsrate des BIP pro Kopf von ost- und westdeutschen Landkreisen



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

```

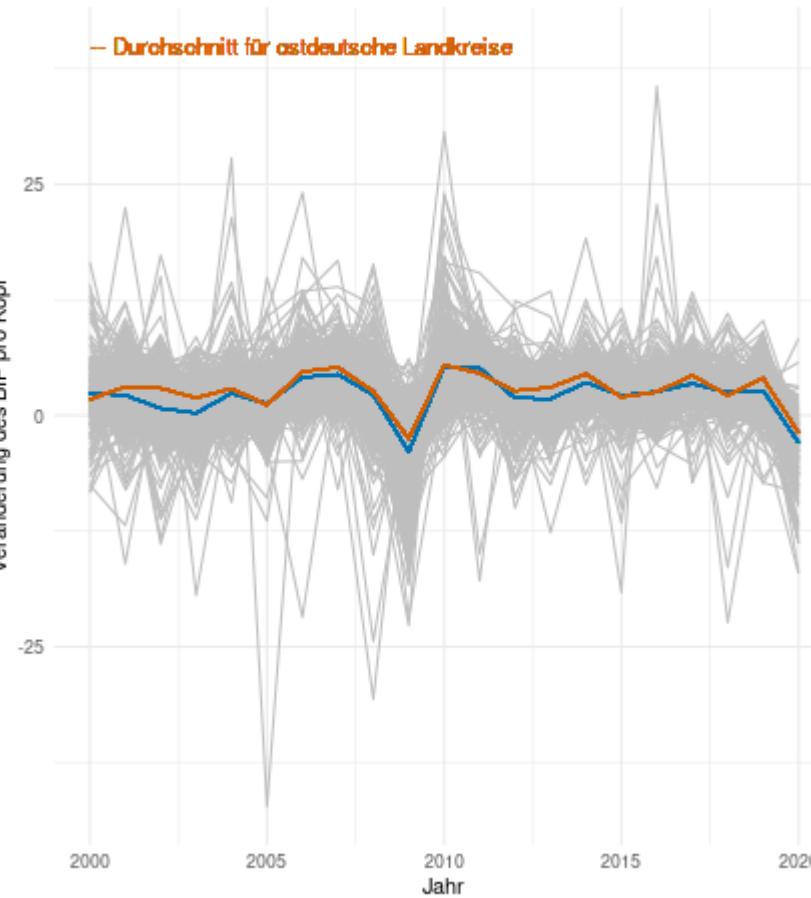
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf_wachstum,
  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 pro Kopf von ost- und westdeutschen
  caption = "Quelle: Daten der Statistischen Ämter der Länder und des Bundes.
  x = "Jahr",
  y = "Veränderung des BIP pro Kopf") +
  theme(legend.position = "none") +
  geom_text(aes(x=2000, y=40, label = "-- Durchschni

```

Die Wachstumsrate des BIP pro Kopf von ost- und westdeutschen Landkreisen

— Durchschnitt für ostdeutsche Landkreise



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

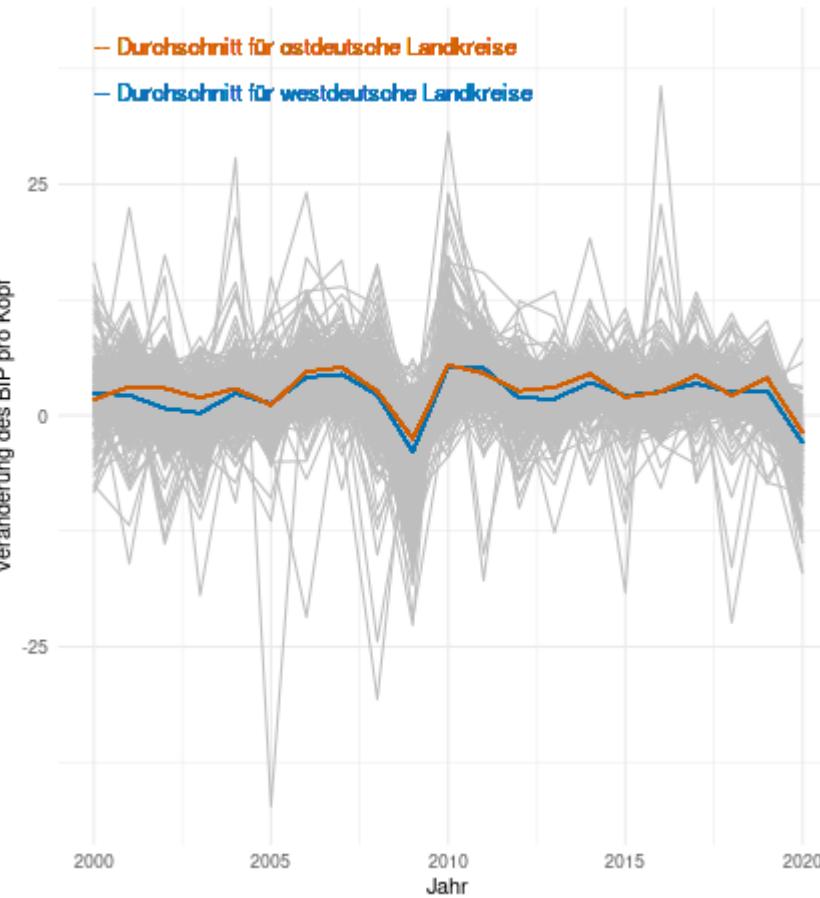
```

bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf_wachstum,
  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 pro Kopf v
  caption = "Quelle: Daten der Statistischen Äm
  x = "Jahr",
  y = "Veränderung des BIP pro Kopf") +
  theme(legend.position = "none") +
  geom_text(aes(x=2000, y=40, label = "-- Durchschni
  geom_text(aes(x=2000, y=35, label = "-- Durchschni

```

Die Wachstumsrate des BIP pro Kopf von ost- und westdeutschen Landkreisen



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

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.

Bruttoinlandsprodukt pro Kopf

Bisherige Grafiken:

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

Bruttoinlandsprodukt pro Kopf

Bisherige Grafiken:

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

Alternative Darstellungen der Verteilung:

- + Histogramm (nächste Folie)
- + Kerndichteschätzer (siehe ausführliche Case-Study)

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 × 14
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002          12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck  01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01    Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01    Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01   Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01   Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01   Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01   Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01   Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

```
gesamtdaten %>%
  group_by(ost_name)

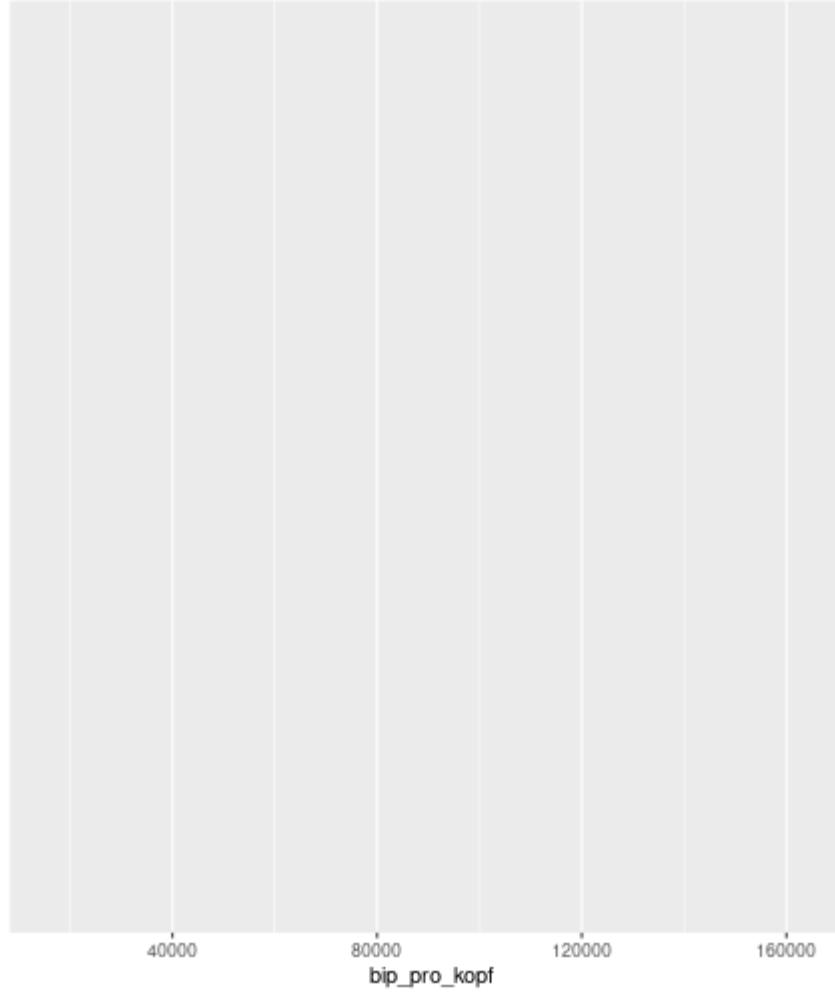
## # A tibble: 401 × 14
## # Groups:   ost_name [2]
##   Regionalsch...¹ total...² landk...³ bunde...⁴ bunde...⁵ Schul...⁶ Einwo...⁷ Schul...⁸   lk
##   <chr>           <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001            4512 Flensb... 01     Schles...  5791.   87770  5.08e8 3.67e
## 2 01002            12345 Kiel     01     Schles...  3839.   247135 9.49e8 1.14e
## 3 01003            9692 Lübeck   01     Schles...  5567.   216739 1.21e9 9.16e
## 4 01004            3836 Neumün... 01     Schles...  5409.   78759  4.26e8 3.34e
## 5 01051            4632 Dithma... 01     Schles...  1670.   133684 2.23e8 4.47e
## 6 01053            5592 Herzog... 01     Schles...  1293.   195677 2.53e8 4.50e
## 7 01054            5657 Nordfr... 01     Schles...  2624.   165642 4.35e8 5.74e
## 8 01055            5748 Osthol... 01     Schles...  1890.   200931 3.80e8 5.27e
## 9 01056            8599 Pinneb... 01     Schles...  2225.   311713 6.94e8 9.07e
## 10 01057           3264 Plön     01     Schles...  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 5 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, and abbreviated variable names
## #   `¹Regionalschlüssel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

```
gesamtdaten %>%  
  group_by(ost_name) %>%  
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm
```

```
## # A tibble: 2 × 2  
##   ost_name      durchschnitt  
##   <chr>          <dbl>  
## 1 Ostdeutschland    28091.  
## 2 Westdeutschland   38884.
```

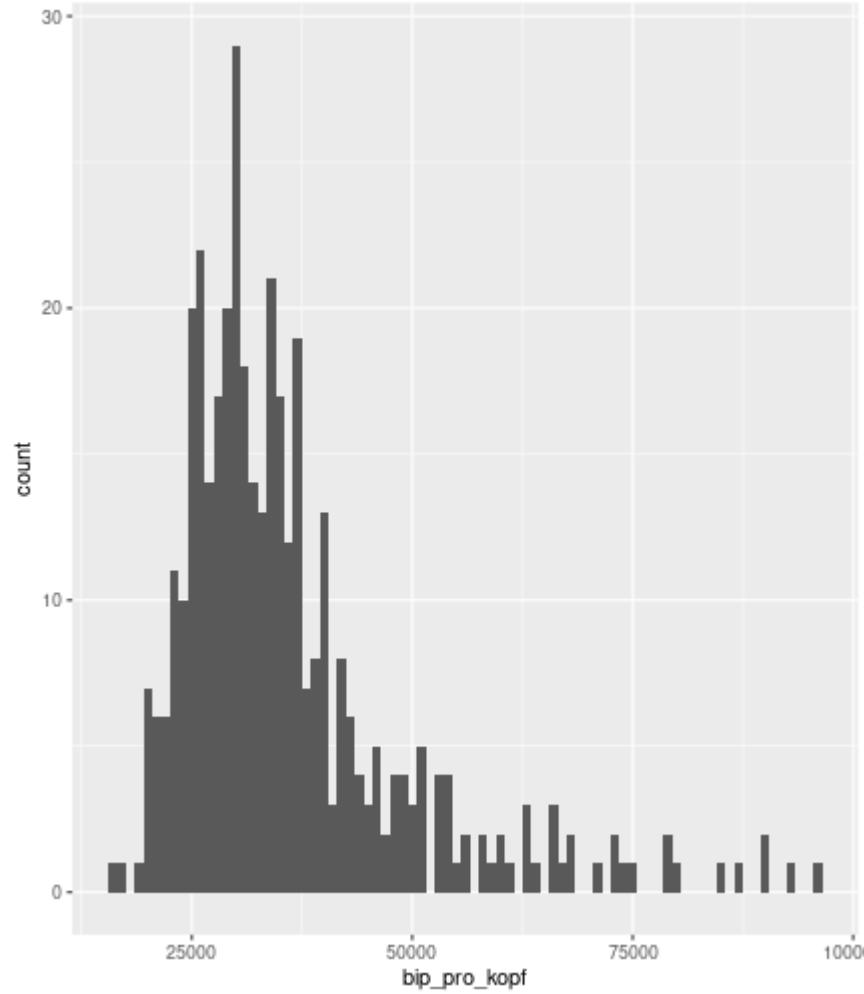
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel
```

```
gesamtdaten %>%  
  group_by(ost_name) %>%  
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm  
  ungroup() -> mittel  
  
ggplot(gesamtdaten, aes(x = bip_pro_kopf))
```



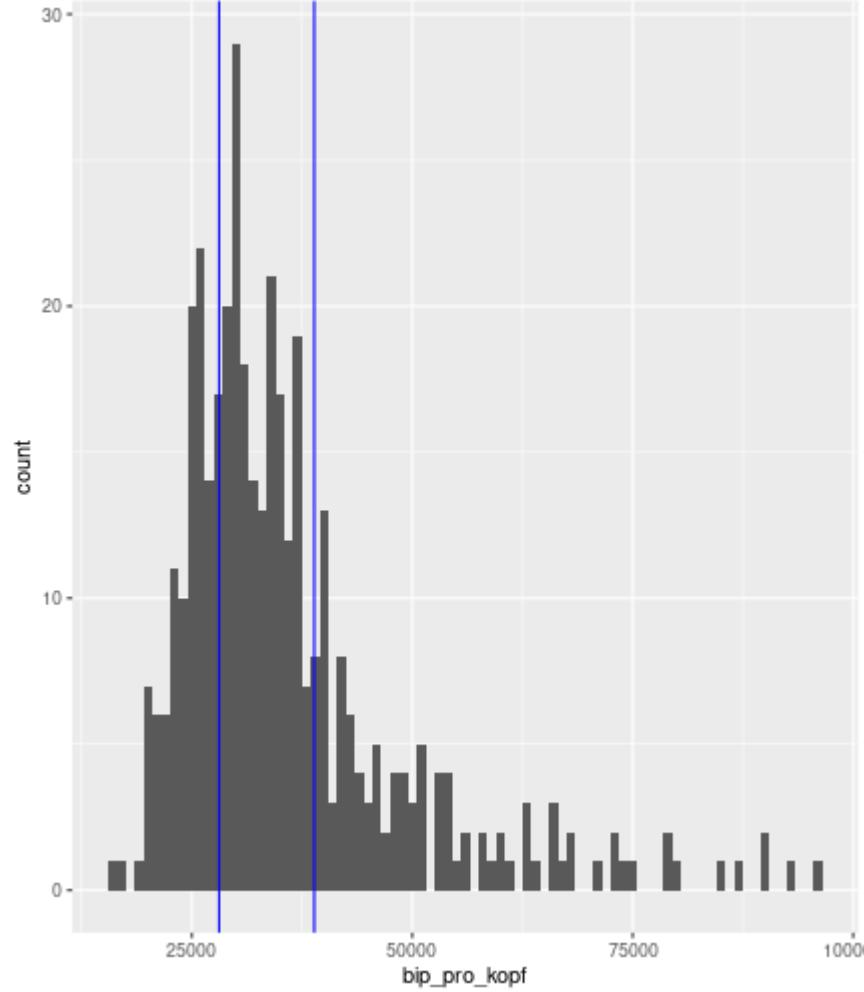
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf >= 0),
```



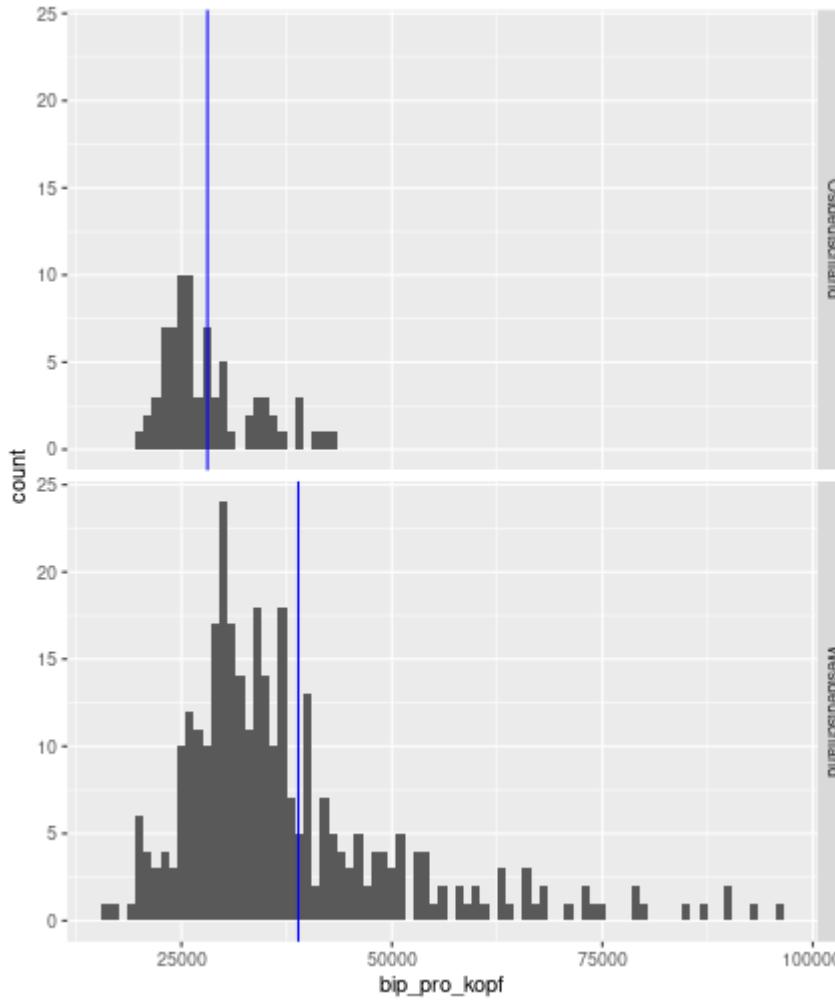
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf >= 0), bins = 20)
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
```



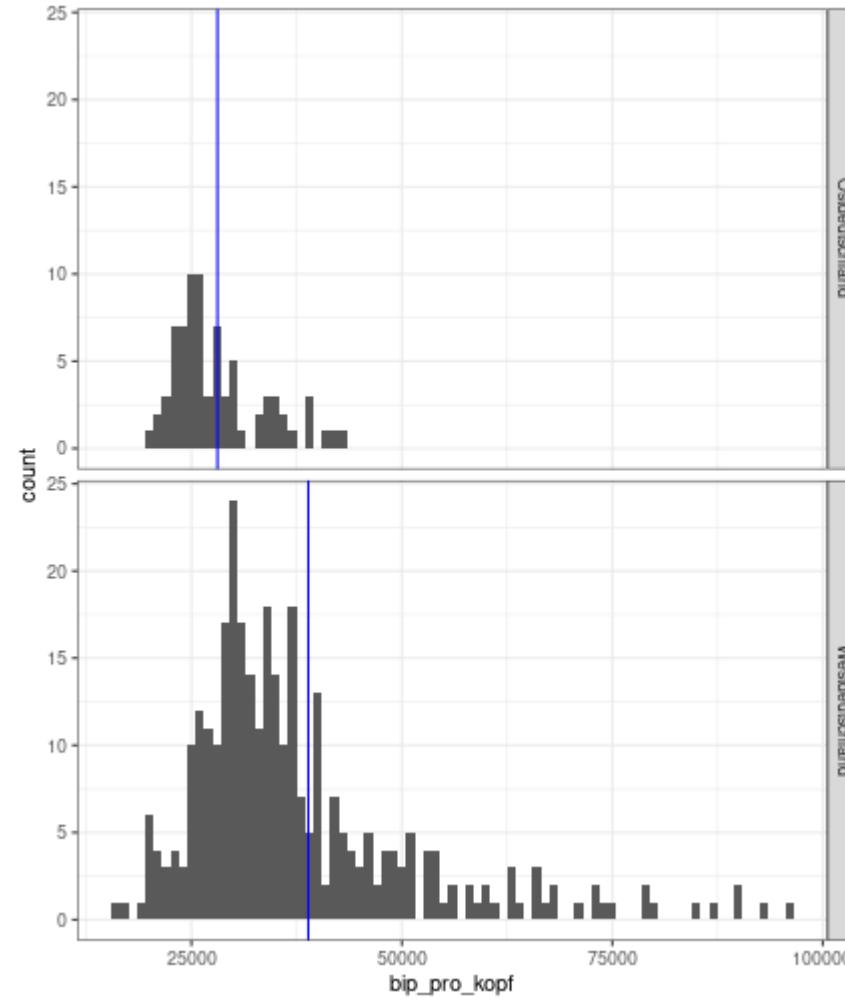
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
ungroup() -> mittel

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



```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, ost_name == "Ostdeutschland"),
  geom_vline(data = mittel, aes(xintercept = durchschnitt),
  facet_grid(ost_name ~ .) +
  theme_bw()
```

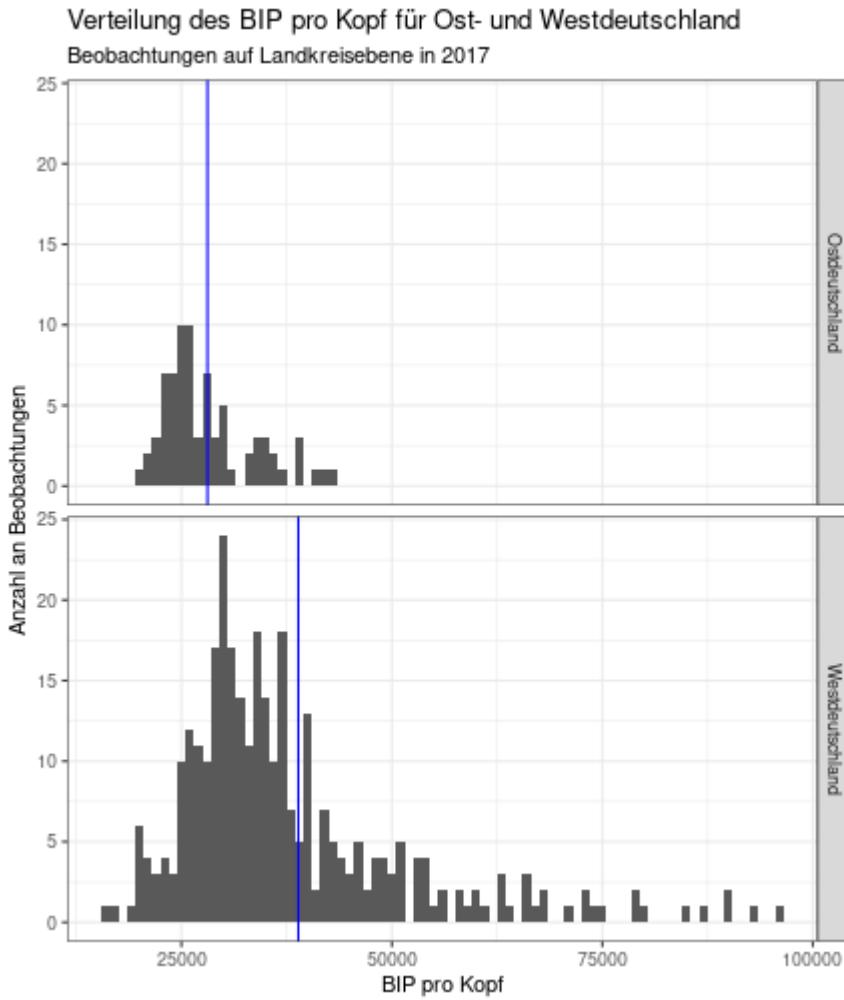


```

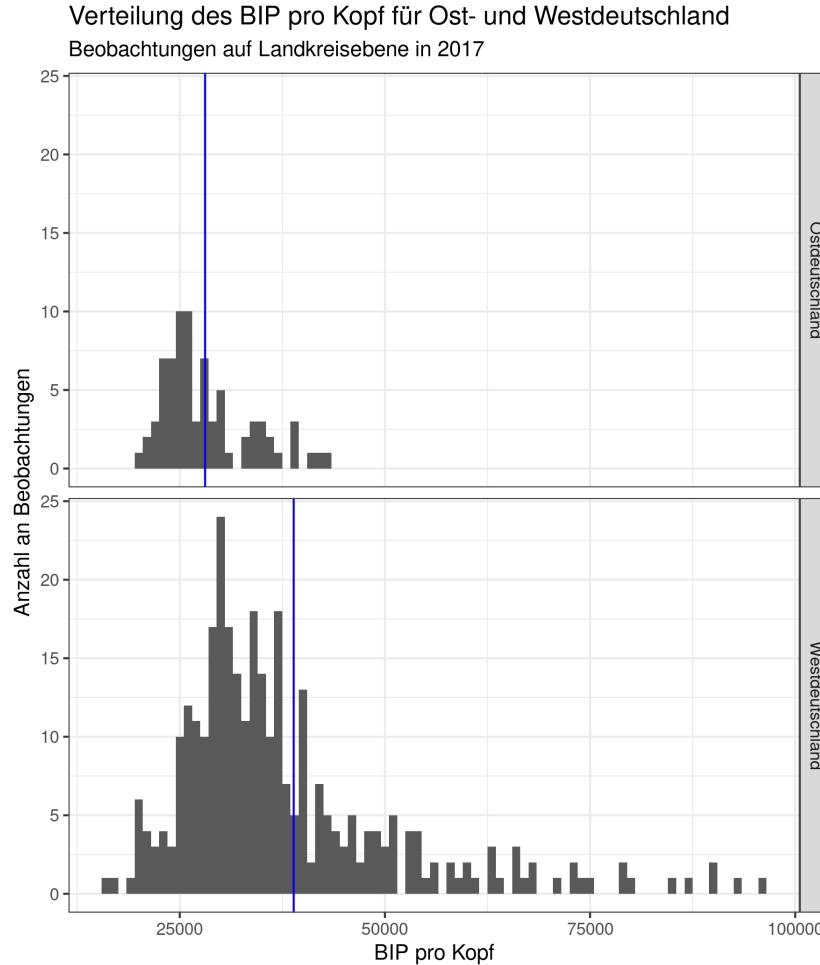
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel

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

```



Verteilung des BIP pro Kopf in 2017



Verteilung des BIP pro Kopf in 2017

Das 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 28091€
- + 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

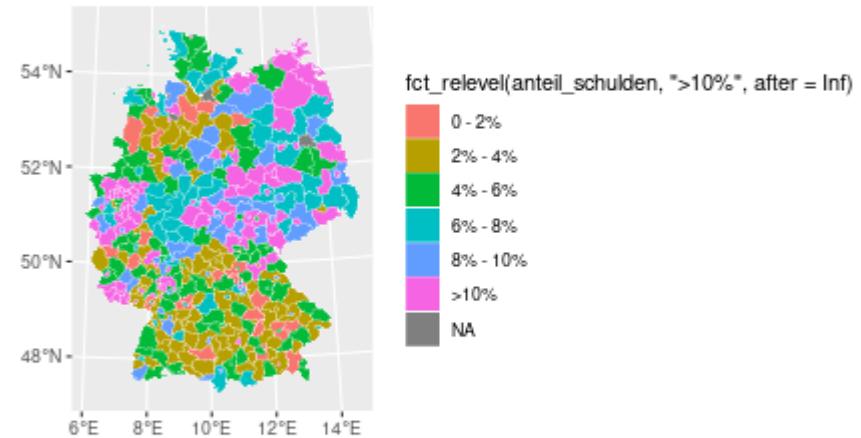
Warum könnte die Verschuldung des öffentlichen Haushalts ein Indikator für eine hohe Arbeitslosenquote sein?

Darstellung der Verschuldung der Landkreise mittels einer Deutschlandkarte.

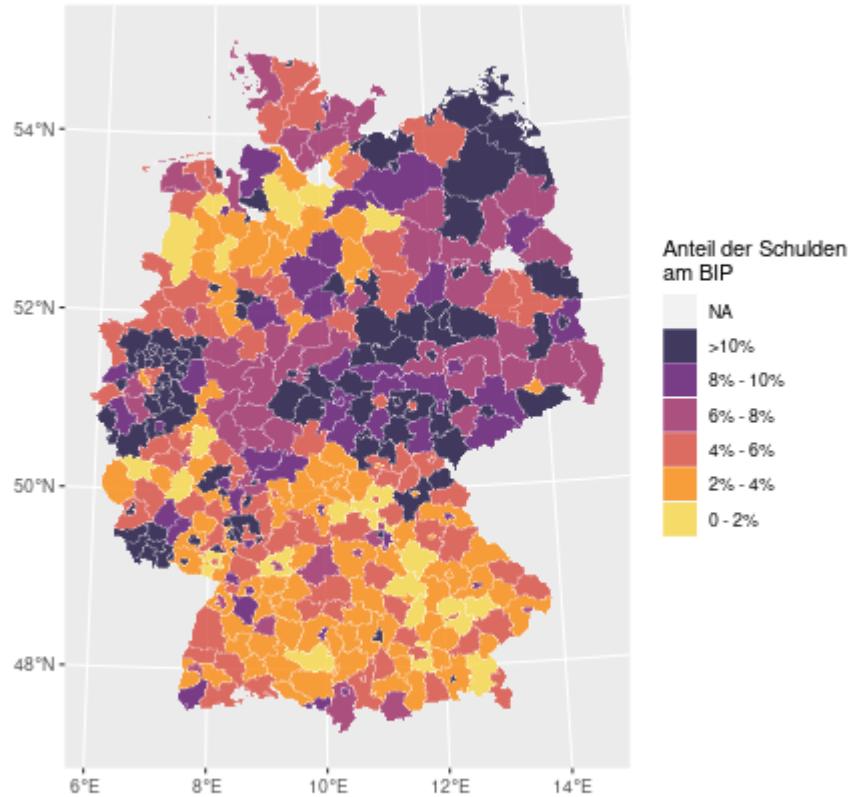
Beschreiben und interpretieren Sie die folgende Grafik.

```
ggplot(  
# Datensatz  
data = schulden_landkreise_anteil  
)
```

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



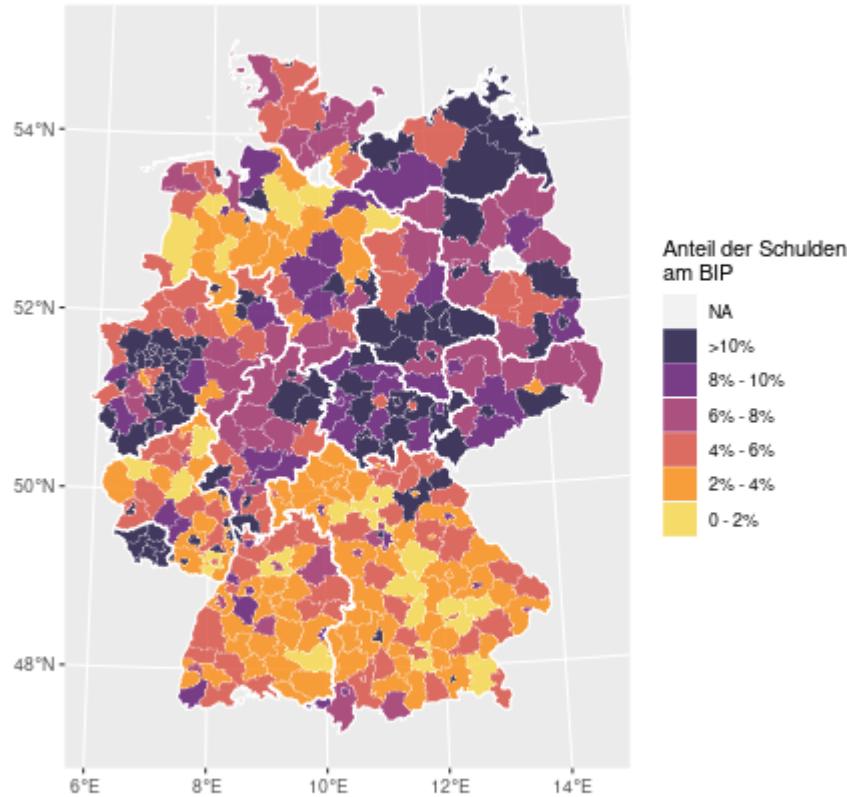
```
ggplot(  
  # Datensatz  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af  
    ),  
    color = "white",  
    size = 0.1  
  ) +  
  # Viridis Farbschema  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "Anteil der Schulden\nam BIP",  
    alpha = 0.8, # Deckkraft der Füllung  
    begin = 0.1,  
    end = 0.9,  
    direction = -1,  
    guide = guide_legend(reverse = T))
```



```

ggplot(
# Datensatz
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af
    ),
    color = "white",
    size = 0.1
  ) +
# Viridis Farbschema
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # Deckkraft der Füllung
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# etwas dickere Linien für Bundeslandgrenzen
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "white",
    size = 0.5
)

```

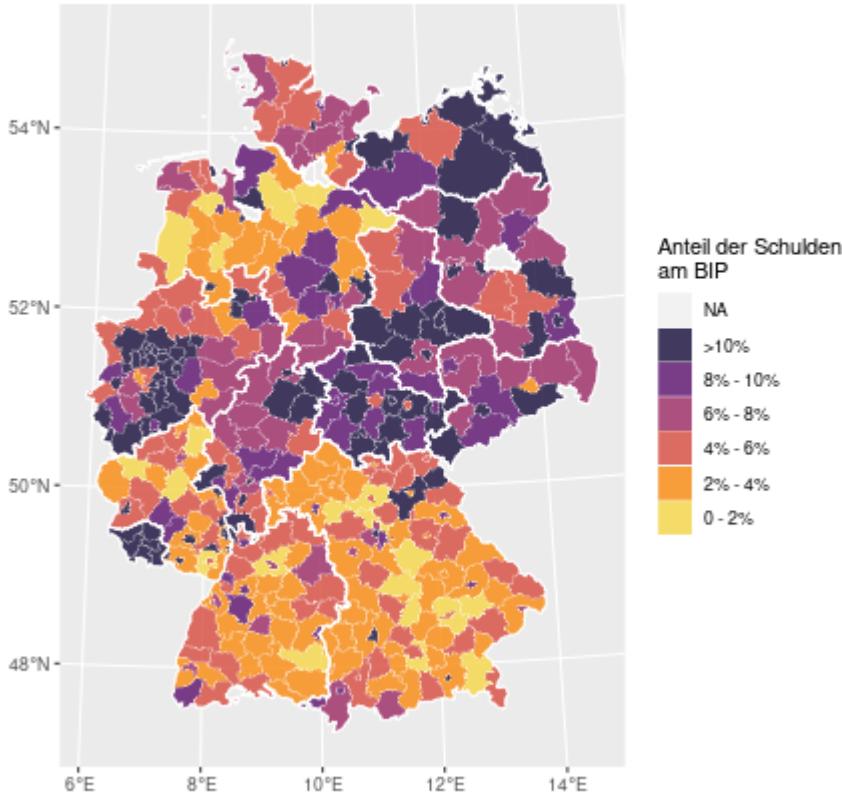


```

ggplot(
# Datensatz
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af
    ),
    color = "white",
    size = 0.1
  ) +
# Viridis Farbschema
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # Deckkraft der Füllung
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# etwas dickere Linien für Bundeslandgrenzen
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "white",
    size = 0.5
  ) +
# Titel
  labs(x = NULL,
       y = NULL,
       title = "Wie verschuldet sind die deutschen Landkreise?",
       subtitle = "Öffentliche Schulden im Vergleich zum BIP in 2017")

```

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



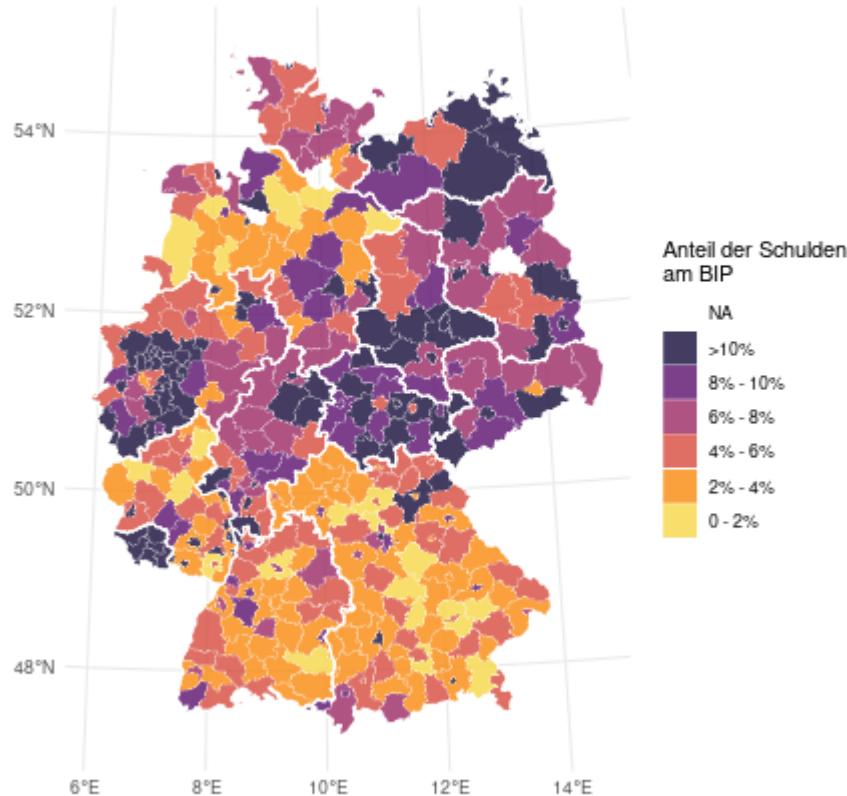
```
ggplot(  
# Datensatz  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af  
    ),  
    color = "white",  
    size = 0.1  
  ) +  
# Viridis Farbschema  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "Anteil der Schulden\nam BIP",  
    alpha = 0.8, # Deckkraft der Füllung  
    begin = 0.1,  
    end = 0.9,  
    direction = -1,  
    guide = guide_legend(reverse = T)) +  
# etwas dickere Linien für Bundeslandgrenzen  
  geom_sf(  
    data = bundesland,  
    fill = "transparent",  
    color = "white",  
    size = 0.5  
  ) +  
# Titel  
  labs(x = NULL,  
       y = NULL,  
       title = "Wie verschuldet sind die deutschen L  
       subtitle = "Öffentliche Schulden im Vergleich  
       theme_minimal() -> plot_schulden_lk
```

```

ggplot(
# Datensatz
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af
    ),
    color = "white",
    size = 0.1
  ) +
# Viridis Farbschema
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # Deckkraft der Füllung
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# etwas dickere Linien für Bundeslandgrenzen
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "white",
    size = 0.5
  ) +
# Titel
  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_1k
)

```

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



Verschuldung

Beschreibung:

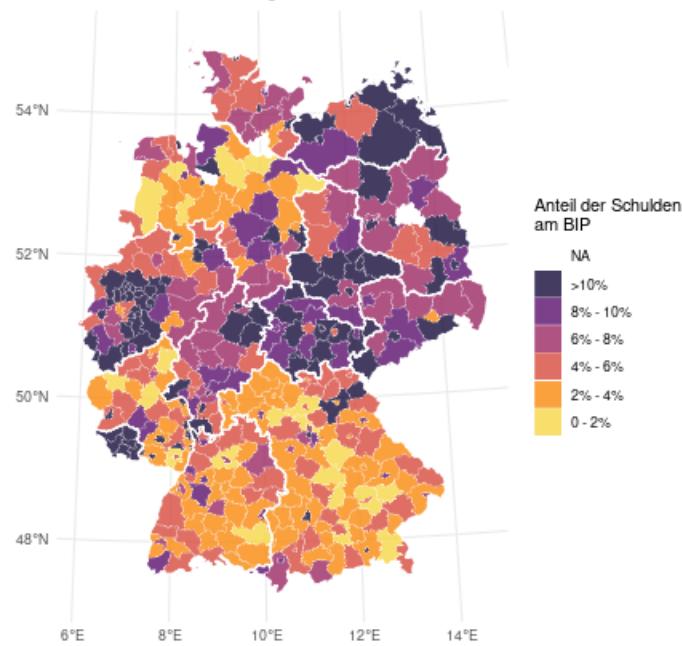
- + Niedrige Verschuldung im Verhältnis zum BIP: Bayern, Baden-Württemberg, Rheinland-Pflanz, Niedersachsen
- + Hohe Verschuldung: Nordrhein-Westfalen, Saarland, Sachse, Mecklenburg-Vorpommern
- + Mittlere Verschuldung: Brandenburg, Thüringen, Hessen

Interpretation:

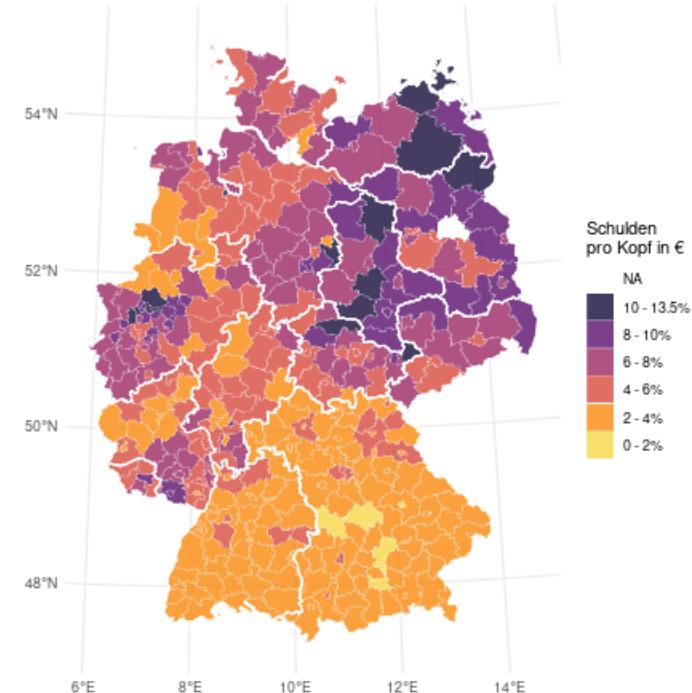
- + Strukturschwache Landkreise sind vermehrt in Ostdeutschland zu finden, allerdings scheint es eher ein Nord/Süd Gefälle als ein Ost/West Gefälle zu geben
- + Die ehemalige Herzammer der deutschen Industrie, das Ruhrgebiet, leidet unter dem Strukturwandel hin zu erneuerbaren Energien
 - + Es fallen 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-pro-Kopf-Wachstum zusammen?

Korrelation zwischen Arbeitslosenquote und BIP-pro-Kopf-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-pro-Kopf-Wachstum
 - + Variable y: Arbeitslosenquote

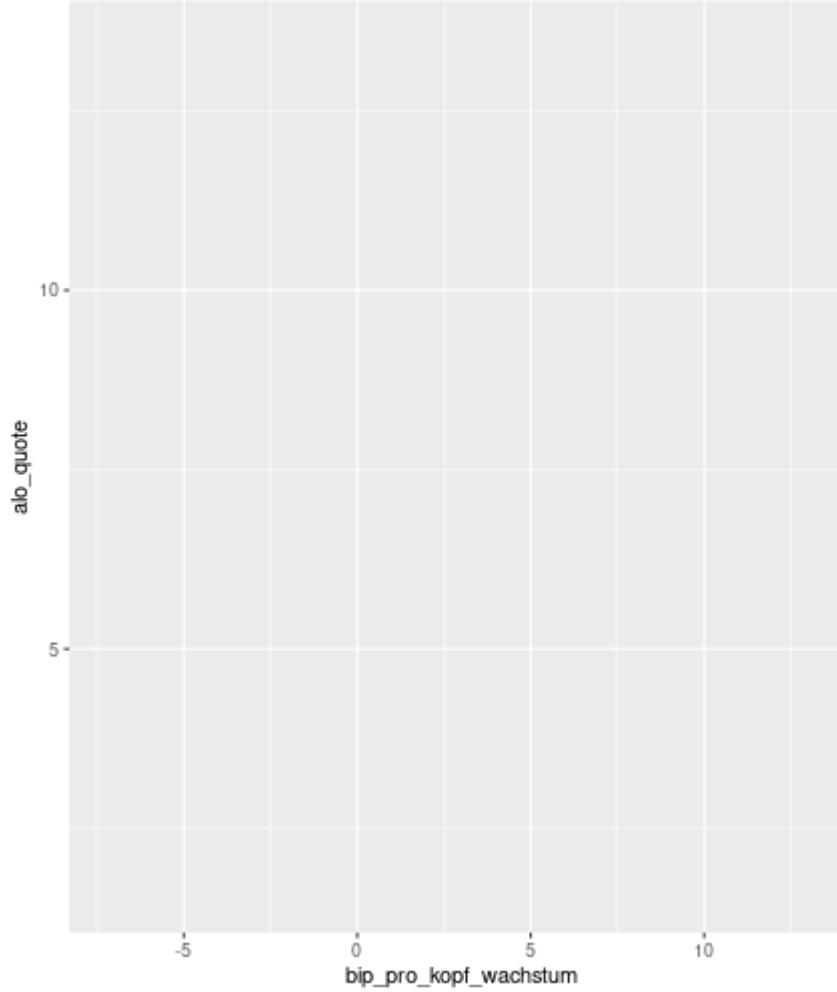
Im Streudiagramm können wir Variable x auf der x-Achse und Variable y auf der y-Achse abtragen

gesamtdaten

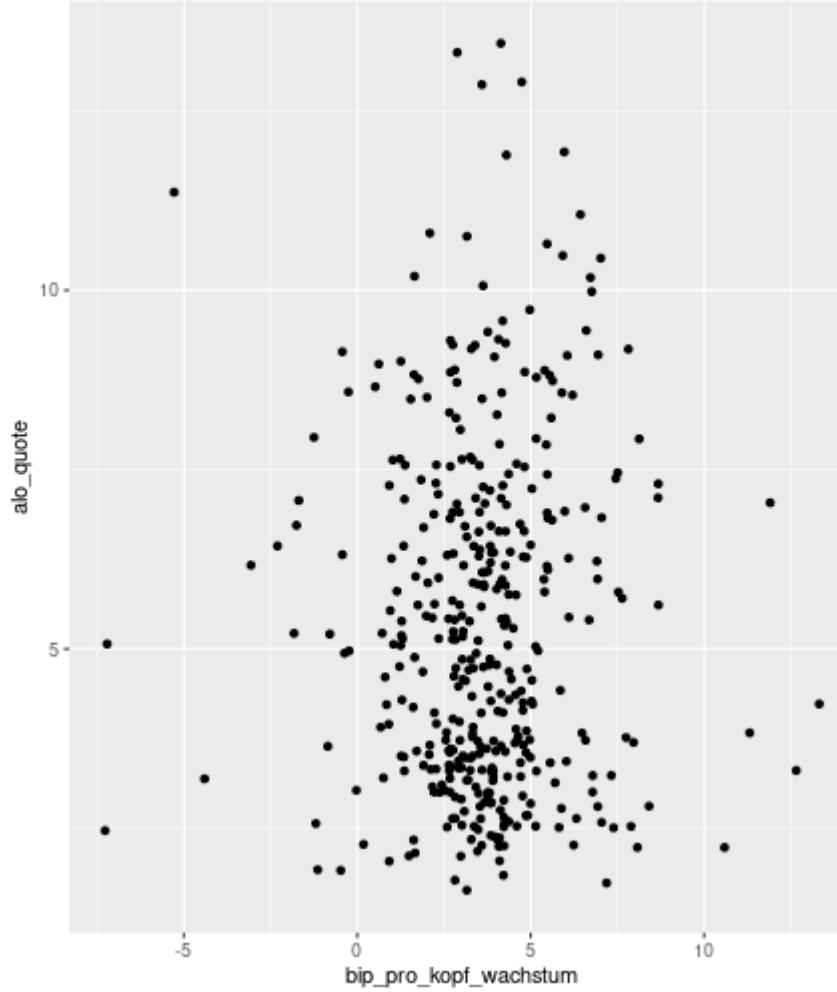
```
## # A tibble: 401 × 17
##   Regionalsch...1 total...2 landk...3 bunde...4 bunde...5 Schul...6 Einwo...7 Schul...8 k
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 01001           4512 Flensb.. 01     Schles..  5791.   87770  5.08e8 3.67e
## 2 01002          12345 Kiel     01     Schles..  3839.   247135 9.49e8 1.14e
## 3 01003           9692 Lübeck  01     Schles..  5567.   216739 1.21e9 9.16e
## 4 01004           3836 Neumün... 01    Schles..  5409.   78759  4.26e8 3.34e
## 5 01051           4632 Dithma... 01    Schles..  1670.   133684 2.23e8 4.47e
## 6 01053           5592 Herzog... 01   Schles..  1293.   195677 2.53e8 4.50e
## 7 01054           5657 Nordfr... 01   Schles..  2624.   165642 4.35e8 5.74e
## 8 01055           5748 Osthol... 01   Schles..  1890.   200931 3.80e8 5.27e
## 9 01056           8599 Pinneb... 01   Schles..  2225.   311713 6.94e8 9.07e
## 10 01057          3264 Plön     01   Schles..  1532.   128763 1.97e8 2.55e
## # ... with 391 more rows, 8 more variables: bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, bip_pro_kopf_wachstum <dbl>,
## #   Jahr <dbl>, anteil_schulden <dbl>, and abbreviated variable names
## #   `¹Regionalschluessel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

```
gesamtdaten %>%
```

```
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot
```

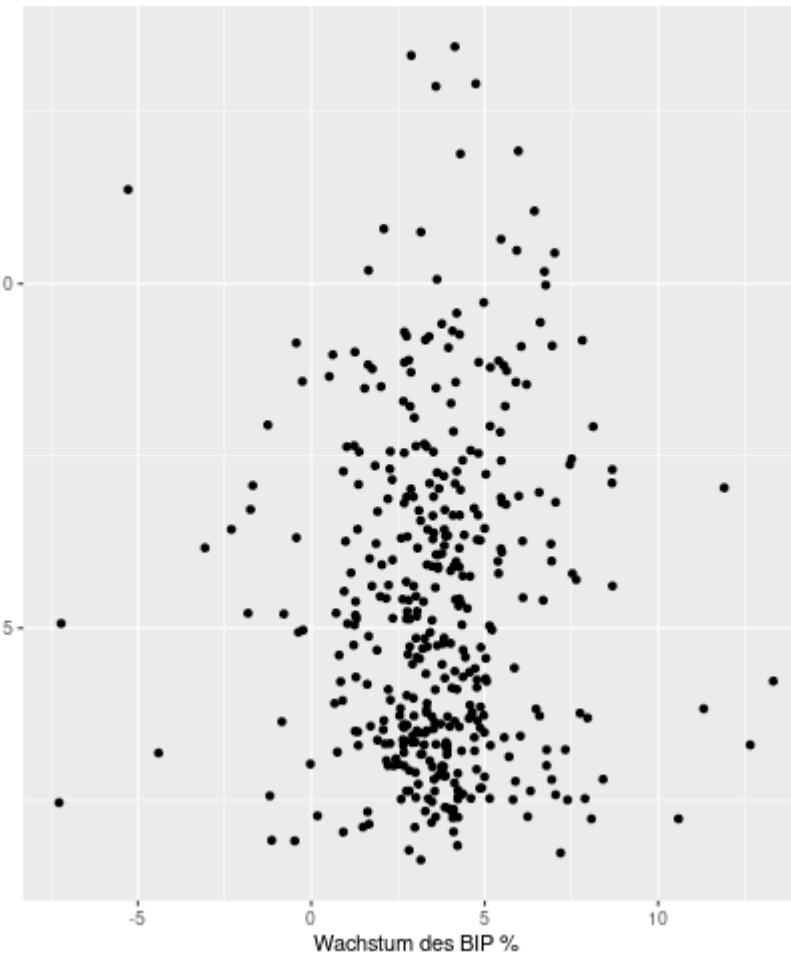


```
gesamtdaten %>%  
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot  
  geom_point()
```



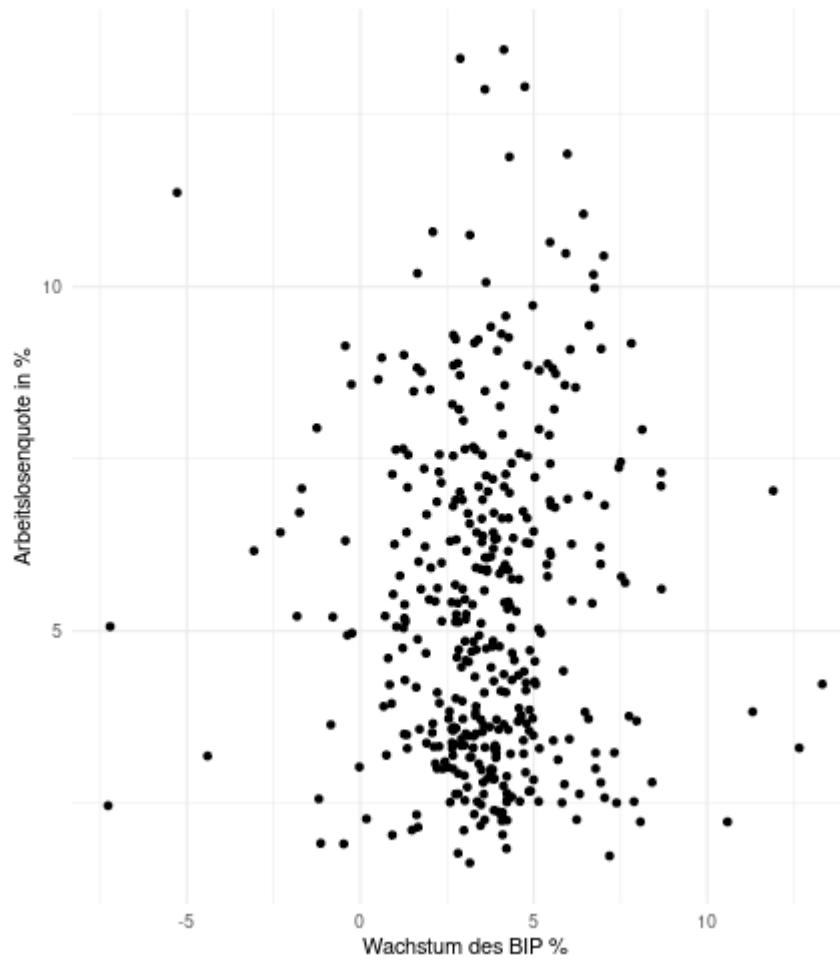
```
gesamtdaten %>%  
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot  
  geom_point() +  
  labs( x = "Wachstum des BIP %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation des BIP-Wachstums und d
```

Korrelation des BIP-Wachstums und der Arbeitslosenquote



```
gesamtdaten %>%  
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot  
  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_pro_kopf_wachstum,  
     use = "pairwise.complete.obs")
```

```
## [1] 0.02202843
```

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_pro_kopf_wachstum,  
     use = "pairwise.complete.obs")
```

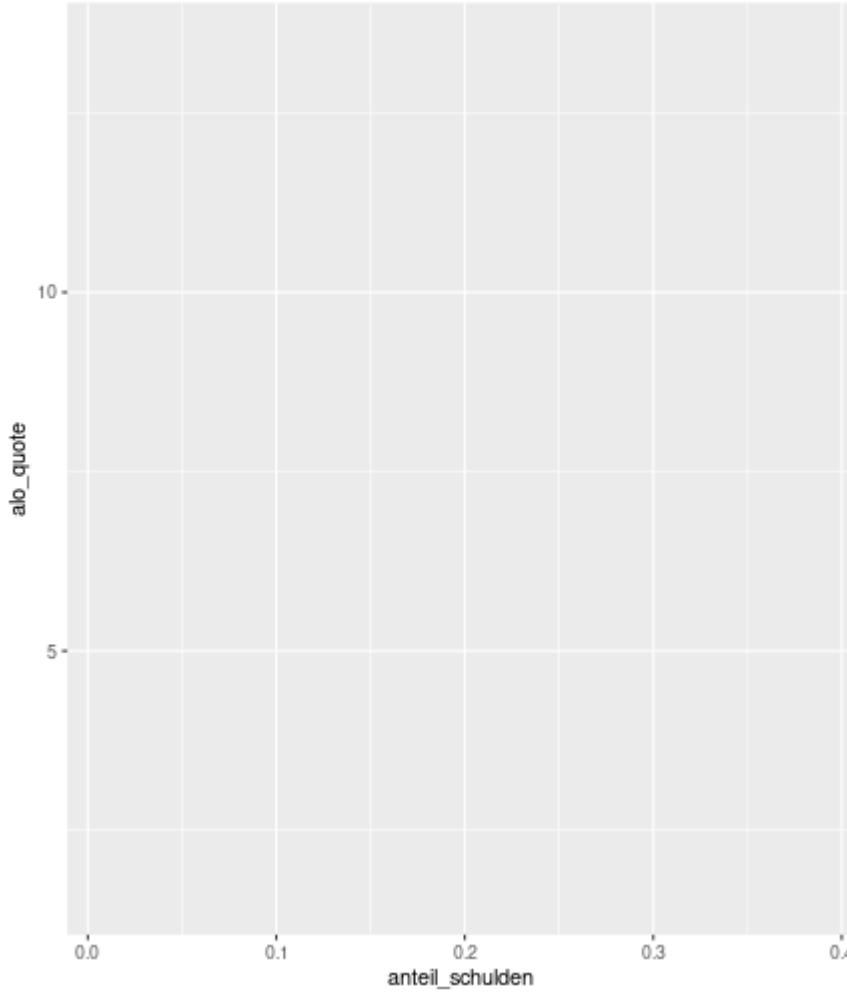
```
## [1] 0.02202843
```

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

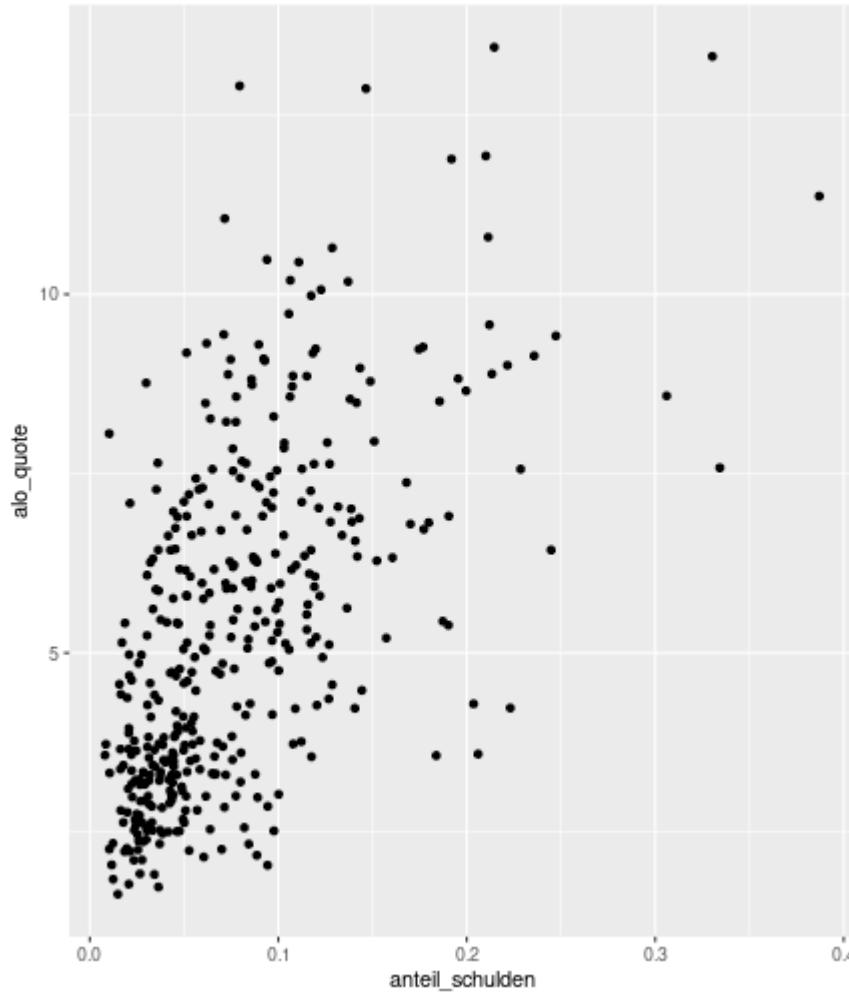
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g
```

```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten  
  
## # A tibble: 401 × 17  
##   Regionalsch...¹ total...² landk...³ bunde...⁴ bunde...⁵ Schul...⁶ Einwo...⁷ Schul...⁸ k  
##   <chr>          <dbl> <chr>    <chr>    <chr>    <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 01001           4512 Flensb.. 01     Schles.. 5791. 87770 5.08e8 3.67e  
## 2 01002           12345 Kiel    01     Schles.. 3839. 247135 9.49e8 1.14e  
## 3 01003           9692 Lübeck  01     Schles.. 5567. 216739 1.21e9 9.16e  
## 4 01004           3836 Neumün... 01    Schles.. 5409. 78759 4.26e8 3.34e  
## 5 01051           4632 Dithma... 01    Schles.. 1670. 133684 2.23e8 4.47e  
## 6 01053           5592 Herzog... 01   Schles.. 1293. 195677 2.53e8 4.50e  
## 7 01054           5657 Nordfr... 01   Schles.. 2624. 165642 4.35e8 5.74e  
## 8 01055           5748 Osthol... 01   Schles.. 1890. 200931 3.80e8 5.27e  
## 9 01056           8599 Pinneb... 01   Schles.. 2225. 311713 6.94e8 9.07e  
## 10 01057          3264 Plön    01   Schles.. 1532. 128763 1.97e8 2.55e  
## # ... with 391 more rows, 8 more variables: bip_pro_kopf <dbl>, erw <dbl>,  
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, bip_pro_kopf_wachstum <dbl>,  
## #   Jahr <dbl>, anteil_schulden <dbl>, and abbreviated variable names  
## #   `¹Regionalschluessel`, `²total_alo`, `³landkreis_name`, `⁴bundesland`,  
## #   `⁵bundesland_name`, `⁶Schulden_pro_kopf_lk`, `⁷Einwohner`, `⁸Schulden_gesamt
```

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

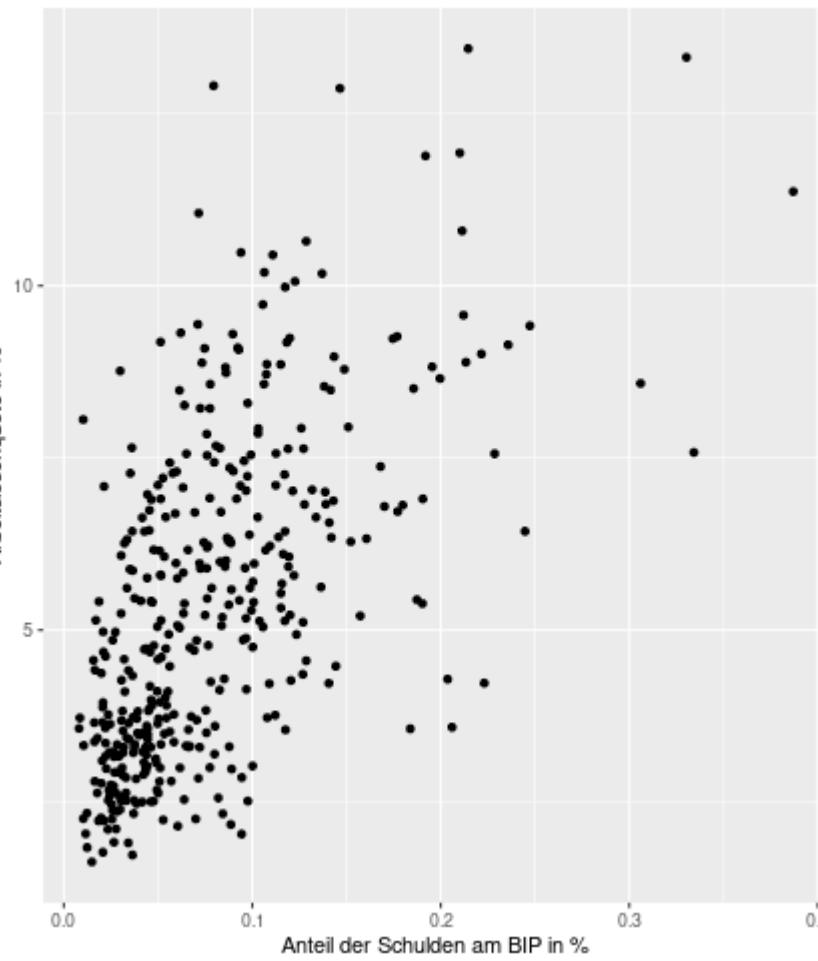


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point()
```

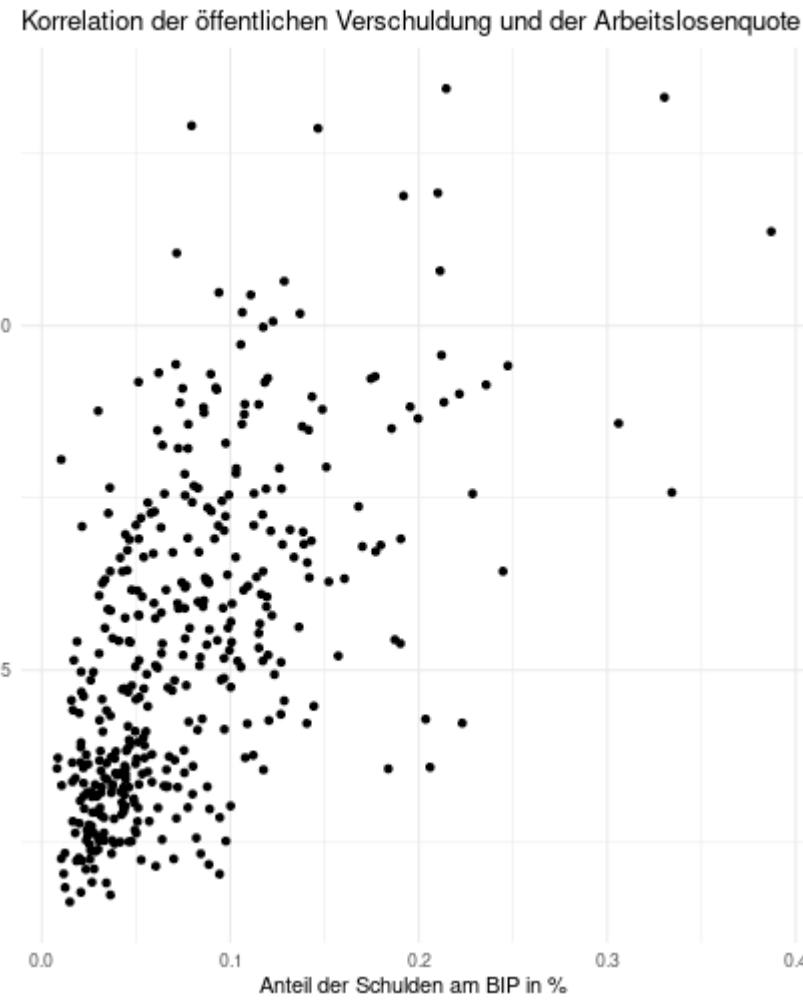


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Anteil der Schulden am BIP in %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation der öffentlichen Versch"
```

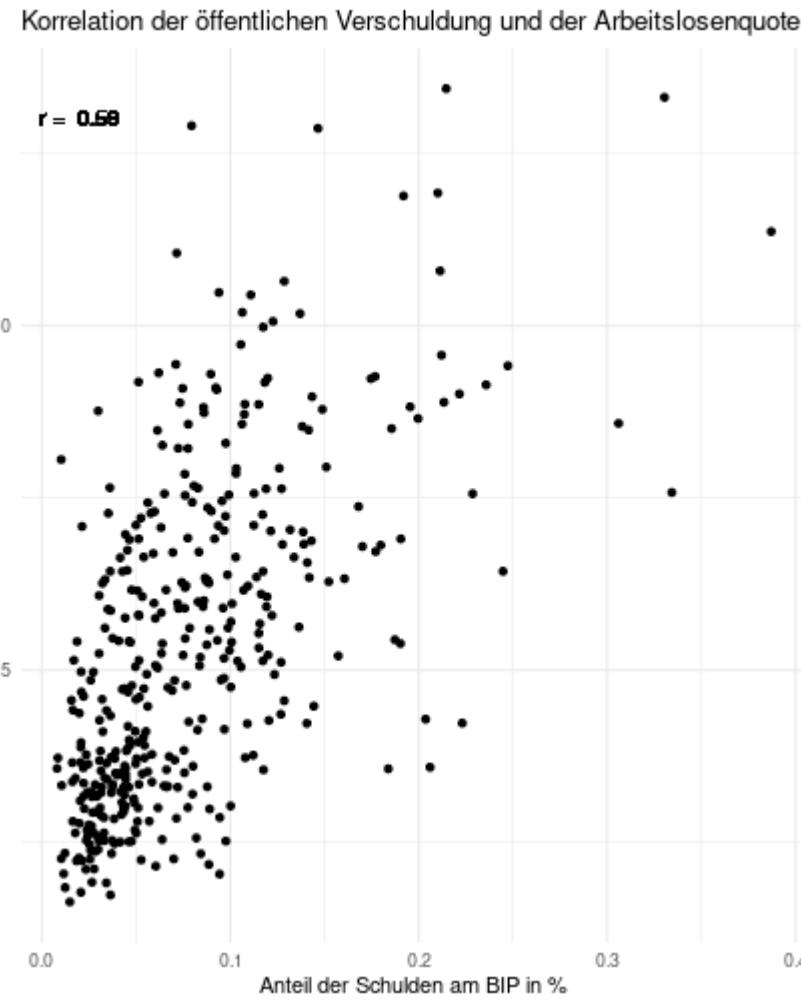
Korrelation der öffentlichen Verschuldung und der Arbeitslosenquote



```
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 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.