

# 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

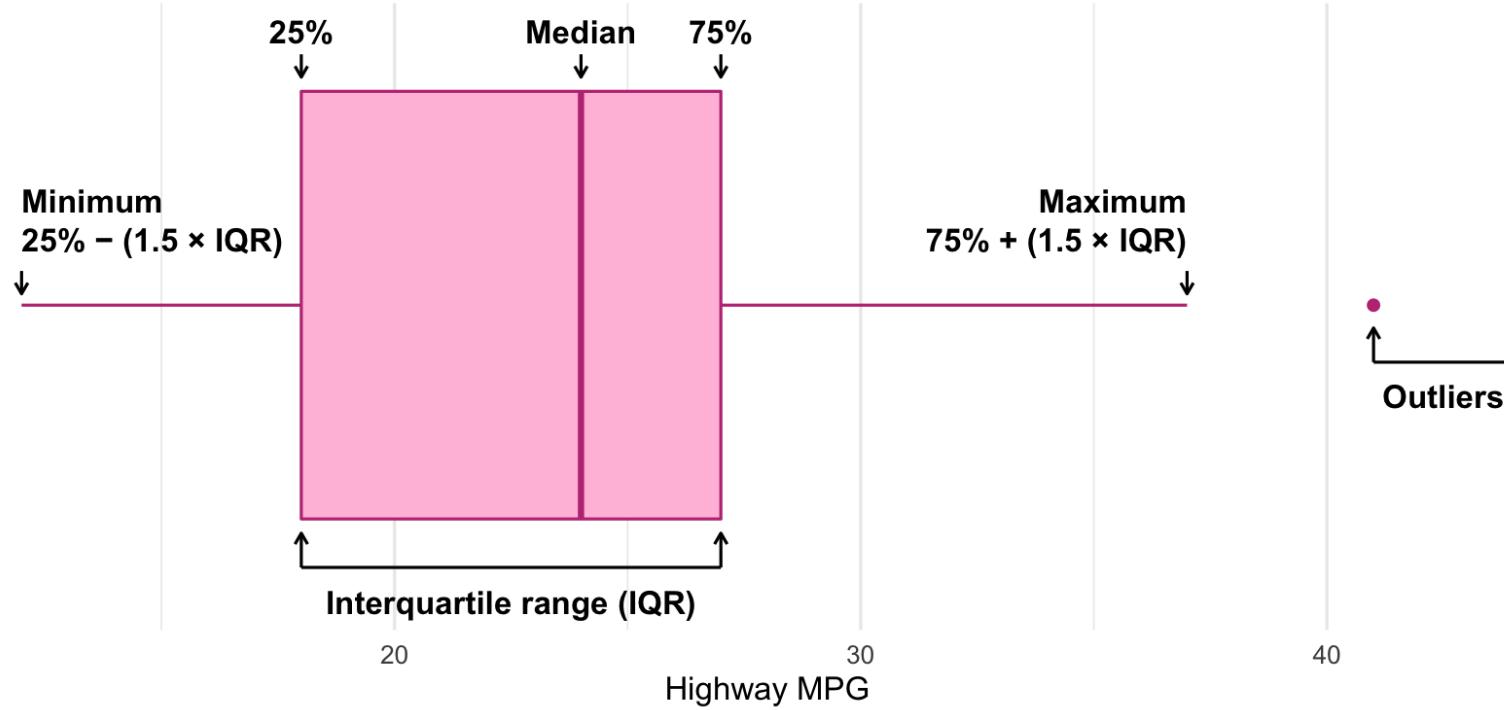
# Univariate deskriptive Statistik

- ✚ 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     41572423        76939741
```

# 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     41572423        76939741
```

- ✚ 41,5 Mio. Erwerbstätige und 76,9 Mio Einwohner in Deutschland
- ✚ Folgende Stadtstaaten sind nicht in unseren Berechnungen enthalten:
  - ✚ Hamburg (1,8 Mio.)
  - ✚ Berlin (3,87 Mio.)
  - ✚ Bremen (0.6 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	400
Number of columns	4

---

## Column type frequency:

character	1
numeric	3

---

Group variables	None
-----------------	------

# Anzahl an Beobachtungen

- ✚ 400 individuelle Beobachtungen in unserem Datensatz.

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

Stimmen diese Angaben?

# Anzahl an Beobachtungen

- ✚ 400 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 106 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	32	0	378	0

- ✚ Nur 378 unterschiedliche Landkreis Namen in unserem Datensatz mit 400 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	32	0	378	0

- ✚ Nur 378 unterschiedliche Landkreis Namen in unserem Datensatz mit 400 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	5.28	2.07	1.81	3.66	5.00	6.62	14.4
Schulden_pro_kopf_lk	4	0.99	2919.82	2272.31	218.94	1462.92	2255.53	3460.74	16678.0
bip_pro_kopf	2	1.00	40241.49	16714.37	17553.39	30303.58	36497.09	43846.03	158749.0
Einwohner	4	0.99	194292.28	151587.20	34029.00	103573.25	155842.00	237861.25	1487560.0

# Anzahl an Beobachtungen

- ✚ Fehlende Beobachtungen für Schulden pro Kopf: *vier* Landkreise
- ✚ Fehlende Beobachtung für Einwohner: *vier* Landkreise
- ✚ Fehlende Beobachtungen für BIP pro Kopf: *zwei* Landkreise
- ✚ Fehlende Beobachtungen für die Arbeitslosenquote: *zwei* 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 Beobachtung für Einwohner: *vier* Landkreise
- ✚ Fehlende Beobachtungen für BIP pro Kopf: *zwei* Landkreise
- ✚ Fehlende Beobachtungen für die Arbeitslosenquote: *zwei* 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	5.28	2.07	1.81	3.66	5.00	6.62	14.47
Schulden_pro_kopf_lk	4	0.99	2919.82	2272.31	218.94	1462.92	2255.53	3460.74	16678.07
bip_pro_kopf	2	1.00	40241.49	16714.37	17553.39	30303.58	36497.09	43846.03	158749.09

# Beschreibung der Tabelle

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	5.28	2.07	1.81	3.66	5.00	6.62	14.47
Schulden_pro_kopf_lk	4	0.99	2919.82	2272.31	218.94	1462.92	2255.53	3460.74	16678.07
bip_pro_kopf	2	1.00	40241.49	16714.37	17553.39	30303.58	36497.09	43846.03	158749.09

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	5.28	2.07	1.81	3.66	5.00	6.62	14.47
Schulden_pro_kopf_lk	4	0.99	2919.82	2272.31	218.94	1462.92	2255.53	3460.74	16678.07
bip_pro_kopf	2	1.00	40241.49	16714.37	17553.39	30303.58	36497.09	43846.03	158749.09

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: 5,28 Prozent

- ✚ Sehr hoch
- ✚ Jedoch SGB II und SGB III
- ✚ Konsistenzcheck auf [Statista](#) zeigt eine Arbeitslosenquote von 5,7% für 2021
- ✚ **Jedoch:** Wir haben nicht Berlin und Hamburg in den Daten

Standardabweichung: 2,07

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

Median: 5,00 Prozent

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

# Verschuldung pro Kopf

Mittelwert: 2920€

- ✚ Moderat von der Höhe her

Standardabweichung: 2272€

- ✚ Sehr hohe Streuung
- ✚ Deutliche regionale Unterschiede

Median: 2256€

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

# BIP pro Kopf

Mittelwert: 40241€

- ✚ Insgesamt recht hoch
- ✚ Starker Wirtschaftsstandort Deutschland

Standardabweichung: 16714€

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

Median: 36497€

- ✚ 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 2021
- ✚ 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 2021 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: 400 × 12
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>           <chr>      <chr>
## 1 01001            4369. Flensburg        01       Schleswig-Holst
## 2 01002            11097. Kiel           01       Schleswig-Holst
## 3 01003            9347. Lübeck         01       Schleswig-Holst
## 4 01004            3771. Neumünster      01       Schleswig-Holst
## 5 01051            4143. Dithmarschen     01       Schleswig-Holst
## 6 01053            5603. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4699. Nordfriesland     01       Schleswig-Holst
## 8 01055            5371. Ostholstein      01       Schleswig-Holst
## 9 01056            9371. Pinneberg        01       Schleswig-Holst
## 10 01057           2854. Plön            01       Schleswig-Holst
## # ... with 390 more rows
## # ... with 7 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>
```

```
gesamtdaten %>%
```

```
  group_by( bundesland_name )
```

```
## # A tibble: 400 × 12
## # Groups:   bundesland_name [16]
##   Regionalschlüssel total_alo landkreis_name   bundesland bundesland_name
##   <chr>           <dbl>   <chr>           <chr>      <chr>
## 1 01001            4369. Flensburg        01       Schleswig-Holst
## 2 01002            11097. Kiel           01       Schleswig-Holst
## 3 01003            9347. Lübeck          01       Schleswig-Holst
## 4 01004            3771. Neumünster      01       Schleswig-Holst
## 5 01051            4143. Dithmarschen     01       Schleswig-Holst
## 6 01053            5603. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4699. Nordfriesland      01       Schleswig-Holst
## 8 01055            5371. Ostholstein       01       Schleswig-Holst
## 9 01056            9371. Pinneberg        01       Schleswig-Holst
## 10 01057           2854. Plön            01       Schleswig-Holst
## # i 390 more rows
## # i 7 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>
```

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

```
## # A tibble: 16 x 4  
##   bundesland_name     mean_alo    sd_alo median_alo  
##   <chr>          <dbl>     <dbl>      <dbl>  
## 1 Baden-Württemberg    3.80     0.795     3.73  
## 2 Bayern                3.31     0.746     3.23  
## 3 Berlin                NA      NA        NA  
## 4 Brandenburg           6.73     1.50      7.03  
## 5 Bremen                9.43     2.32      9.43  
## 6 Hamburg                NA      NA        NA  
## 7 Hessen                5.11     1.38      5.18  
## 8 Mecklenburg-Vorpommern 7.55     1.18      7.67  
## 9 Niedersachsen          5.86     1.73      5.90  
## 10 Nordrhein-Westfalen   7.14     2.46      6.85  
## 11 Rheinland-Pfalz       5.53     1.53      5.39  
## 12 Saarland               6.08     1.82      5.75  
## 13 Sachsen                5.80     0.893     5.58  
## 14 Sachsen-Anhalt         7.70     1.33      7.47  
## 15 Schleswig-Holstein    5.92     0.915     5.95  
## 16 Thüringen              6.10     1.36      5.57
```

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

# Die Arbeitslosenquote auf Bundeslandebene

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

```
## # A tibble: 14 x 4
##   bundesland_name   mean_alo   sd_alo   median
##   <chr>          <dbl>     <dbl>     <
## 1 Bayern           3.31      0.746
## 2 Baden-Württemberg 3.80      0.795
## 3 Hessen           5.11      1.38
## 4 Rheinland-Pfalz  5.53      1.53
## 5 Sachsen          5.80      0.893
## 6 Niedersachsen   5.86      1.73
## 7 Schleswig-Holstein 5.92      0.915
## 8 Saarland         6.08      1.82
## 9 Thüringen        6.10      1.36
## 10 Brandenburg    6.73      1.50
## 11 Nordrhein-Westfalen 7.14      2.46
## 12 Mecklenburg-Vorpommern 7.55      1.18
## 13 Sachsen-Anhalt  7.70      1.33
## 14 Bremen           9.43      2.32
```

# 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>    <dbl>
## 1 Bayern           3.31     0.746   3.31
## 2 Baden-Württemberg 3.80     0.795   3.80
## 3 Hessen           5.11     1.38    5.11
## 4 Rheinland-Pfalz  5.53     1.53    5.53
## 5 Sachsen          5.80     0.893   5.80
## 6 Niedersachsen    5.86     1.73    5.86
## 7 Schleswig-Holstein 5.92     0.915   5.92
## 8 Saarland          6.08     1.82    6.08
## 9 Thüringen         6.10     1.36    6.10
## 10 Brandenburg      6.73     1.50    6.73
## 11 Nordrhein-Westfalen 7.14     2.46    7.14
## 12 Mecklenburg-Vorpommern 7.55     1.18    7.55
## 13 Sachsen-Anhalt   7.70     1.33    7.70
## 14 Bremen            9.43     2.32    9.43
```

Bundesland	Arbeitslosenquote		
	Mittelwert	Std.	Median
Bayern	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

Bitte beachten:

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

<sup>1</sup> 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(5,7,9, 10,12,13), 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  3.80    0.795     3.73
## 2 Bayern            3.31    0.746     3.23
## 3 Berlin             NA     NA        NA
## 4 Brandenburg       6.73    1.50      7.03
## 5 Bremen             9.43    2.32      9.43
## 6 Hamburg            NA     NA        NA
## 7 Hessen              5.11    1.38      5.18
## 8 Mecklenburg-Vorpommern 7.55    1.18      7.67
## 9 Niedersachsen      5.86    1.73      5.90
## 10 Nordrhein-Westfalen 7.14    2.46      6.85
## 11 Rheinland-Pfalz   5.53    1.53      5.39
## 12 Saarland           6.08    1.82      5.75
## 13 Sachsen            5.80    0.893     5.58
## 14 Sachsen-Anhalt    7.70    1.33      7.47
## 15 Schleswig-Holstein 5.92    0.915     5.95
## 16 Thüringen          6.10    1.36      5.57
```

```
bula_data %>%  
  arrange( mean_alo )
```

```
## # A tibble: 16 x 4  
##   bundesland_name      mean_alo   sd_alo median_alo  
##   <chr>          <dbl>    <dbl>     <dbl>  
## 1 Bayern           3.31     0.746    3.23  
## 2 Baden-Württemberg 3.80     0.795    3.73  
## 3 Hessen           5.11     1.38     5.18  
## 4 Rheinland-Pfalz  5.53     1.53     5.39  
## 5 Sachsen          5.80     0.893    5.58  
## 6 Niedersachsen   5.86     1.73     5.90  
## 7 Schleswig-Holstein 5.92     0.915    5.95  
## 8 Saarland          6.08     1.82     5.75  
## 9 Thüringen         6.10     1.36     5.57  
## 10 Brandenburg     6.73     1.50     7.03  
## 11 Nordrhein-Westfalen 7.14     2.46     6.85  
## 12 Mecklenburg-Vorpommern 7.55     1.18     7.67  
## 13 Sachsen-Anhalt   7.70     1.33     7.47  
## 14 Bremen            9.43     2.32     9.43  
## 15 Berlin            NA      NA       NA  
## 16 Hamburg           NA      NA       NA
```

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

```
## # A tibble: 14 x 4  
##   bundesland_name      mean_alo   sd_alo median_alo  
##   <chr>          <dbl>    <dbl>      <dbl>  
## 1 Bayern            3.31     0.746     3.23  
## 2 Baden-Württemberg 3.80     0.795     3.73  
## 3 Hessen            5.11     1.38      5.18  
## 4 Rheinland-Pfalz  5.53     1.53      5.39  
## 5 Sachsen           5.80     0.893     5.58  
## 6 Niedersachsen    5.86     1.73      5.90  
## 7 Schleswig-Holstein 5.92     0.915     5.95  
## 8 Saarland          6.08     1.82      5.75  
## 9 Thüringen         6.10     1.36      5.57  
## 10 Brandenburg      6.73     1.50      7.03  
## 11 Nordrhein-Westfalen 7.14     2.46      6.85  
## 12 Mecklenburg-Vorpommern 7.55     1.18      7.67  
## 13 Sachsen-Anhalt   7.70     1.33      7.47  
## 14 Bremen            9.43     2.32      9.43
```

```
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	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

```

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	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

```

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	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

```

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(5, 7, 9, 10, 12, 13), bold = T, color = "wh

```

Bundesland	Mittelwert	Std.	Median
Bayern	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

```

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(5, 7, 9, 10, 12, 13), bold = T, color = "wh
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Arbeitslosenquote				
Bundesland	Mittelwert	Std.	Median	
Bayern	3.31	0.75	3.23	
Baden-Württemberg	3.80	0.80	3.73	
Hessen	5.11	1.38	5.18	
Rheinland-Pfalz	5.53	1.53	5.39	
Sachsen	5.80	0.89	5.58	
Niedersachsen	5.86	1.73	5.90	
Schleswig-Holstein	5.92	0.92	5.95	
Saarland	6.08	1.82	5.75	
Thüringen	6.10	1.36	5.57	
Brandenburg	6.73	1.50	7.03	
Nordrhein-Westfalen	7.14	2.46	6.85	
Mecklenburg-Vorpommern	7.55	1.18	7.67	
Sachsen-Anhalt	7.70	1.33	7.47	
Bremen	9.43	2.32	9.43	

```

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(5, 7, 9, 10, 12, 13), bold = T, color = "wh
  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	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

*Bitte beachten:*

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

<sup>1</sup> Die ostdeutschen Bundesländer sind grau hinterlegt.

Arbeitslosenquote

Bundesland	Mittelwert	Std.	Median
Bayern	3.31	0.75	3.23
Baden-Württemberg	3.80	0.80	3.73
Hessen	5.11	1.38	5.18
Rheinland-Pfalz	5.53	1.53	5.39
Sachsen	5.80	0.89	5.58
Niedersachsen	5.86	1.73	5.90
Schleswig-Holstein	5.92	0.92	5.95
Saarland	6.08	1.82	5.75
Thüringen	6.10	1.36	5.57
Brandenburg	6.73	1.50	7.03
Nordrhein-Westfalen	7.14	2.46	6.85
Mecklenburg-Vorpommern	7.55	1.18	7.67
Sachsen-Anhalt	7.70	1.33	7.47
Bremen	9.43	2.32	9.43

*Bitte beachten:*

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

<sup>1</sup> 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"))
```

gesamtdata

```
## # A tibble: 400 × 12
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>           <chr>      <chr>
## 1 01001            4369. Flensburg            01       Schleswig-Holst
## 2 01002            11097. Kiel                01       Schleswig-Holst
## 3 01003            9347. Lübeck              01       Schleswig-Holst
## 4 01004            3771. Neumünster          01       Schleswig-Holst
## 5 01051            4143. Dithmarschen         01       Schleswig-Holst
## 6 01053            5603. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4699. Nordfriesland        01       Schleswig-Holst
## 8 01055            5371. Ostholstein          01       Schleswig-Holst
## 9 01056            9371. Pinneberg            01       Schleswig-Holst
## 10 01057           2854. Plön                01       Schleswig-Holst
## # i 390 more rows
## # i 7 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>
```

```
gesamtdaten %>%
```

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

```
## # A tibble: 400 × 14
##   Regionalschluessel total_alo landkreis_name      bundesland bundesland_name
##   <chr>                <dbl> <chr>                <chr>      <chr>
## 1 01001                 4369. Flensburg            01        Schleswig-Holst
## 2 01002                 11097. Kiel                01        Schleswig-Holst
## 3 01003                 9347. Lübeck              01        Schleswig-Holst
## 4 01004                 3771. Neumünster        01        Schleswig-Holst
## 5 01051                 4143. Dithmarschen        01        Schleswig-Holst
## 6 01053                 5603. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                 4699 Nordfriesland        01        Schleswig-Holst
## 8 01055                 5371 Ostholstein          01        Schleswig-Holst
## 9 01056                 9371. Pinneberg           01        Schleswig-Holst
## 10 01057                2854. Plön                01        Schleswig-Holst
## # i 390 more rows
## # i 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>
```

# Die Arbeitslosenquote zwischen Ost- und Westdeutschland

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

gesamtdaten

```
## # A tibble: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  4369. Flensburg          01        Schleswig-Holst
## 2 01002                  11097. Kiel              01        Schleswig-Holst
## 3 01003                  9347. Lübeck            01        Schleswig-Holst
## 4 01004                  3771. Neumünster        01        Schleswig-Holst
## 5 01051                  4143. Dithmarschen        01        Schleswig-Holst
## 6 01053                  5603. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4699 Nordfriesland        01        Schleswig-Holst
## 8 01055                  5371 Ostholstein          01        Schleswig-Holst
## 9 01056                  9371. Pinneberg          01        Schleswig-Holst
## 10 01057                 2854. Plön               01        Schleswig-Holst
## # ... i 390 more rows
## # ... i 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #       Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #       alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  group_by(ost_name)
```

```
## # A tibble: 400 x 14
## # Groups:   ost_name [2]
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>           <chr>      <chr>
## 1 01001            4369. Flensburg            01       Schleswig-Holst
## 2 01002            11097. Kiel                01       Schleswig-Holst
## 3 01003            9347. Lübeck              01       Schleswig-Holst
## 4 01004            3771. Neumünster          01       Schleswig-Holst
## 5 01051            4143. Dithmarschen        01       Schleswig-Holst
## 6 01053            5603. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4699 Nordfriesland        01       Schleswig-Holst
## 8 01055            5371 Ostholstein          01       Schleswig-Holst
## 9 01056            9371. Pinneberg           01       Schleswig-Holst
## 10 01057           2854. Plön                01       Schleswig-Holst
## # ... with 390 more rows
## # ... with 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>
```

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

```
## # A tibble: 2 × 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 6.65    1.46    4.19    5.44    6.63    7.60    11.0
## 2 Westdeutschland 4.96    2.07    1.81    3.45    4.33    6.27    14.5
```

```
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 6.65    1.46    4.19    5.44    6.63    7.60    11.0  
## 2 Westdeutschland 4.96    2.07    1.81    3.45    4.33    6.27    14.5
```

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

Bundesland	Mittelwert	Std.	Minimum	P25	Median	P75	Maximum
Ostdeutschland	6.65	1.46	4.19	5.44	6.63	7.60	10.98
Westdeutschland	4.96	2.07	1.81	3.45	4.33	6.27	14.47

```

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	6.65	1.46	4.19	5.44	6.63	7.60	10.98
Westdeutschland	4.96	2.07	1.81	3.45	4.33	6.27	14.47

```

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	6.65	1.46	4.19	5.44	6.63	7.60	10.98
Westdeutschland	4.96	2.07	1.81	3.45	4.33	6.27	14.47

```

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	6.65	1.46	4.19	5.44	6.63	7.60	10.98
Westdeutschland	4.96	2.07	1.81	3.45	4.33	6.27	14.47

```

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	6.65	1.46	4.19	5.44	6.63	7.60	10.98
Westdeutschland	4.96	2.07	1.81	3.45	4.33	6.27	14.47

*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:

- ✚ Mittelwert über 1,5 Prozentpunkte niedriger in den Landkreisen der westdeutschen Bundesländer
- ✚ Die Standardabweichung ist in Westdeutschland deutlich höher als in Ostdeutschland
- ✚ Der Median der ostdeutschen Landkreise liegt nahe dem Mittelwert dieser Landkreise. Der Median in den westdeutschen Landkreisen liegt jedoch deutlich unter deren Mittelwert
- ✚ Im 25% Quantil in den **ostdeutschen Landkreisen** ist die Arbeitslosenquote bei **5,44%**
- ✚ Bei den **westdeutschen Landkreisen** ist das **75% Quantil** bei einer Arbeitslosenquote von **6,27%**!

# 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(5,7,9, 10,12,13), 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: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  4369. Flensburg          01        Schleswig-Holst
## 2 01002                  11097. Kiel              01        Schleswig-Holst
## 3 01003                  9347. Lübeck            01        Schleswig-Holst
## 4 01004                  3771. Neumünster        01        Schleswig-Holst
## 5 01051                  4143. Dithmarschen        01        Schleswig-Holst
## 6 01053                  5603. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4699 Nordfriesland        01        Schleswig-Holst
## 8 01055                  5371 Ostholstein          01        Schleswig-Holst
## 9 01056                  9371. Pinneberg          01        Schleswig-Holst
## 10 01057                 2854. Plön               01        Schleswig-Holst
## # ... i 390 more rows
## # ... i 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #       Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #       alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  group_by( bundesland_name )
```

```
## # A tibble: 400 × 14
## # Groups:   bundesland_name [16]
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>           <chr>      <chr>
## 1 01001            4369. Flensburg            01       Schleswig-Holst
## 2 01002            11097. Kiel                01       Schleswig-Holst
## 3 01003            9347. Lübeck              01       Schleswig-Holst
## 4 01004            3771. Neumünster          01       Schleswig-Holst
## 5 01051            4143. Dithmarschen         01       Schleswig-Holst
## 6 01053            5603. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4699. Nordfriesland        01       Schleswig-Holst
## 8 01055            5371. Ostholstein          01       Schleswig-Holst
## 9 01056            9371. Pinneberg            01       Schleswig-Holst
## 10 01057           2854. Plön                 01       Schleswig-Holst
## # ... i 390 more rows
## # ... i 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #       Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #       alo_quote <dbl>, ost <fct>, ost_name <chr>
```

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

```
## # A tibble: 16 x 7
##   bundesland_name  mean_alo  sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_ko
##   <chr>          <dbl>    <dbl>      <dbl>      <dbl>      <dbl>
## 1 Baden-Württembe...  3.80    0.795    47186.    13467.    294
## 2 Bayern            3.31    0.746    46407.    20302.    209
## 3 Berlin            NA     NA        NA        NA        NA
## 4 Brandenburg       6.73    1.50     32525.    6473.     250
## 5 Bremen            9.43    2.32     45840.    12707.    111
## 6 Hamburg           NA     NA        NA        NA        NA
## 7 Hessen             5.11    1.38     43132.    17906.    370
## 8 Mecklenburg-Vor...  7.55    1.18     31891.    6650.     369
## 9 Niedersachsen     5.86    1.73     38128.    21921.    239
## 10 Nordrhein-Westf...  7.14    2.46     39209.    11811.    428
## 11 Rheinland-Pfalz  5.53    1.53     38135.    17557.    316
## 12 Saarland          6.08    1.82     34172.    8198.     522
## 13 Sachsen           5.80    0.893    31943.    5389.     227
## 14 Sachsen-Anhalt   7.70    1.33     30806.    4641.     281
## 15 Schleswig-Holst...  5.92    0.915    36412.    8320.     300
## 16 Thüringen         6.10    1.36     30304.    6108.     270
## # i 1 more variable: sd_schulden <dbl>
```

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

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

bula_data_all
```

## # A tibble: 16 × 7

bundesland_name	mean_alo	sd_alo	mean_bip_kopf	sd_bip_kopf	mean_schulden_ko	sd_schulden
Baden-Württembe...	3.80	0.795	47186.	13467.	294	294
Bayern	3.31	0.746	46407.	20302.	209	209
Berlin	NA	NA	NA	NA	NA	NA
Brandenburg	6.73	1.50	32525.	6473.	250	250
Bremen	9.43	2.32	45840.	12707.	NA	NA
Hamburg	NA	NA	NA	NA	NA	NA
Hessen	5.11	1.38	43132.	17906.	370	370
Mecklenburg-Vor...	7.55	1.18	31891.	6650.	369	369
Niedersachsen	5.86	1.73	38128.	21921.	239	239
Nordrhein-Westf...	7.14	2.46	39209.	11811.	428	428
Rheinland-Pfalz	5.53	1.53	38135.	17557.	316	316
Saarland	6.08	1.82	34172.	8198.	522	522
Sachsen	5.80	0.893	31943.	5389.	227	227
Sachsen-Anhalt	7.70	1.33	30806.	4641.	281	281
Schleswig-Holst...	5.92	0.915	36412.	8320.	300	300
Thüringen	6.10	1.36	30304.	6108.	270	270

## # i 1 more variable: sd\_schulden <dbl>

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

bula_data_all %>%
  arrange( mean_alo )
```

```
## # A tibble: 16 x 7
##   bundesland_name  mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_ko
##   <chr>          <dbl>    <dbl>      <dbl>      <dbl>      <dbl>
## 1 Bayern          3.31     0.746     46407.     20302.     209
## 2 Baden-Württembe... 3.80     0.795     47186.     13467.     294
## 3 Hessen          5.11     1.38      43132.     17906.     370
## 4 Rheinland-Pfalz 5.53     1.53      38135.     17557.     316
## 5 Sachsen          5.80     0.893     31943.     5389.      227
## 6 Niedersachsen   5.86     1.73      38128.     21921.     239
## 7 Schleswig-Holst... 5.92     0.915     36412.     8320.      300
## 8 Saarland         6.08     1.82      34172.     8198.      522
## 9 Thüringen        6.10     1.36      30304.     6108.      270
## 10 Brandenburg    6.73     1.50      32525.     6473.      250
## 11 Nordrhein-Westf... 7.14     2.46      39209.     11811.     428
## 12 Mecklenburg-Vor... 7.55     1.18      31891.     6650.      369
## 13 Sachsen-Anhalt  7.70     1.33      30806.     4641.      283
## 14 Bremen          9.43     2.32      45840.     12707.     11
## 15 Berlin           NA      NA          NA        NA        NA
## 16 Hamburg          NA      NA          NA        NA        NA
## # i 1 more variable: sd_schulden <dbl>
```

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

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

```
## # A tibble: 13 x 7
##   bundesland_name  mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_kopf
##   <chr>          <dbl>    <dbl>      <dbl>      <dbl>          <dbl>
## 1 Bayern           3.31     0.746     46407.     20302.        20900.
## 2 Baden-Württembe... 3.80     0.795     47186.     13467.        29400.
## 3 Hessen            5.11     1.38      43132.     17906.        37000.
## 4 Rheinland-Pfalz  5.53     1.53      38135.     17557.        31000.
## 5 Sachsen           5.80     0.893     31943.      5389.        22000.
## 6 Niedersachsen    5.86     1.73      38128.     21921.        23000.
## 7 Schleswig-Holst... 5.92     0.915     36412.      8320.        30000.
## 8 Saarland          6.08     1.82      34172.      8198.        52000.
## 9 Thüringen         6.10     1.36      30304.      6108.        27000.
## 10 Brandenburg      6.73     1.50      32525.      6473.        25000.
## 11 Nordrhein-Westf... 7.14     2.46      39209.     11811.        42800.
## 12 Mecklenburg-Vor... 7.55     1.18      31891.      6650.        36000.
## 13 Sachsen-Anhalt   7.70     1.33      30806.      4641.        28000.
## # i 1 more variable: sd_schulden <dbl>
```

```

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

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

```

Bundesland	Deskriptive Tabelle komplett				
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert
Bayern	3.31	0.75	46406.88	20301.64	2098.78
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47
Hessen	5.11	1.38	43132.29	17906.49	3708.42
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35
Sachsen	5.80	0.89	31943.24	5389.00	2278.76
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42
Saarland	6.08	1.82	34172.16	8197.88	5220.69
Thüringen	6.10	1.36	30304.08	6108.02	2702.98
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85

```

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	3.31	0.75	46406.88	20301.64	2098.78	1524.02
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47	2287.43
Hessen	5.11	1.38	43132.29	17906.49	3708.42	3336.00
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35	3657.82
Sachsen	5.80	0.89	31943.24	5389.00	2278.76	734.65
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06	1873.50
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42	1403.58
Saarland	6.08	1.82	34172.16	8197.88	5220.69	1246.74
Thüringen	6.10	1.36	30304.08	6108.02	2702.98	592.37
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85	1333.10
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31	2506.46
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52	1634.25
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85	1448.58

Deskriptive Tabelle komplett

```
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)
```

Bundesland	Mittelwert	Std.	Mittelwert		Std.	Mittelwert		Std.
			Mittelwert	Std.		Mittelwert	Std.	
Bayern	3.31	0.75	46406.88	20301.64	2098.78	1524.02		
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47	2287.43		
Hessen	5.11	1.38	43132.29	17906.49	3708.42	3336.00		
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35	3657.82		
Sachsen	5.80	0.89	31943.24	5389.00	2278.76	734.65		
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06	1873.50		
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42	1403.58		
Saarland	6.08	1.82	34172.16	8197.88	5220.69	1246.74		
Thüringen	6.10	1.36	30304.08	6108.02	2702.98	592.37		
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85	1333.10		
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31	2506.46		
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52	1634.25		
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85	1448.58		

```

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(5, 7, 9, 10, 12, 13), bold = T, color = "wh
  
```

Deskriptive Tabelle komplett

Bundesland	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.31	0.75	46406.88	20301.64	2098.78	1524.02
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47	2287.43
Hessen	5.11	1.38	43132.29	17906.49	3708.42	3336.00
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35	3657.82
Sachsen	5.80	0.89	31943.24	5389.00	2278.76	734.65
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06	1873.50
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42	1403.58
Saarland	6.08	1.82	34172.16	8197.88	5220.69	1246.74
Thüringen	6.10	1.36	30304.08	6108.02	2702.98	592.37
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85	1333.10
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31	2506.46
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52	1634.25
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85	1448.58

```

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(5,7,9, 10,12,13), bold = T, color = "wh
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Deskriptive Tabelle komplett

Bundesland	Arbeitslosenquote		BIP pro Kopf		Schulden pro Kopf	
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	3.31	0.75	46406.88	20301.64	2098.78	1524.02
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47	2287.43
Hessen	5.11	1.38	43132.29	17906.49	3708.42	3336.00
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35	3657.82
Sachsen	5.80	0.89	31943.24	5389.00	2278.76	734.65
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06	1873.50
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42	1403.58
Saarland	6.08	1.82	34172.16	8197.88	5220.69	1246.74
Thüringen	6.10	1.36	30304.08	6108.02	2702.98	592.37
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85	1333.10
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31	2506.46
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52	1634.25
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85	1448.58

```

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(5, 7, 9, 10, 12, 13), bold = T, color = "wh
  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	3.31	0.75	46406.88	20301.64	2098.78	1524.02
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47	2287.43
Hessen	5.11	1.38	43132.29	17906.49	3708.42	3336.00
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35	3657.82
Sachsen	5.80	0.89	31943.24	5389.00	2278.76	734.65
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06	1873.50
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42	1403.58
Saarland	6.08	1.82	34172.16	8197.88	5220.69	1246.74
Thüringen	6.10	1.36	30304.08	6108.02	2702.98	592.37
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85	1333.10
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31	2506.46
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52	1634.25
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85	1448.58

Bitte beachten:

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

<sup>1</sup> 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	3.31	0.75	46406.88	20301.64	2098.78	1524.02
Baden-Württemberg	3.80	0.80	47186.49	13467.27	2944.47	2287.43
Hessen	5.11	1.38	43132.29	17906.49	3708.42	3336.00
Rheinland-Pfalz	5.53	1.53	38134.60	17557.18	3166.35	3657.82
Sachsen	5.80	0.89	31943.24	5389.00	2278.76	734.65
Niedersachsen	5.86	1.73	38128.10	21920.77	2394.06	1873.50
Schleswig-Holstein	5.92	0.92	36412.39	8320.48	3003.42	1403.58
Saarland	6.08	1.82	34172.16	8197.88	5220.69	1246.74
Thüringen	6.10	1.36	30304.08	6108.02	2702.98	592.37
Brandenburg	6.73	1.50	32524.72	6472.87	2506.85	1333.10
Nordrhein-Westfalen	7.14	2.46	39209.31	11811.28	4287.31	2506.46
Mecklenburg-Vorpommern	7.55	1.18	31891.43	6650.10	3695.52	1634.25
Sachsen-Anhalt	7.70	1.33	30805.55	4641.38	2834.85	1448.58

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  $< 33000\text{€}$
- ✚ Westdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf  $> 33000\text{€}$
- ✚ 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  $< 33000\text{€}$
- ✚ Westdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf  $> 33000\text{€}$
- ✚ 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 2021

# 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 2021

Die folgende Grafik sollte enthalten:

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

gesamtdaten

```
## # A tibble: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  4369. Flensburg          01        Schleswig-Holst
## 2 01002                  11097. Kiel              01        Schleswig-Holst
## 3 01003                  9347. Lübeck            01        Schleswig-Holst
## 4 01004                  3771. Neumünster        01        Schleswig-Holst
## 5 01051                  4143. Dithmarschen        01        Schleswig-Holst
## 6 01053                  5603. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4699 Nordfriesland        01        Schleswig-Holst
## 8 01055                  5371 Ostholstein          01        Schleswig-Holst
## 9 01056                  9371. Pinneberg          01        Schleswig-Holst
## 10 01057                 2854. Plön               01        Schleswig-Holst
## # ... i 390 more rows
## # ... i 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #       Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #       alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

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

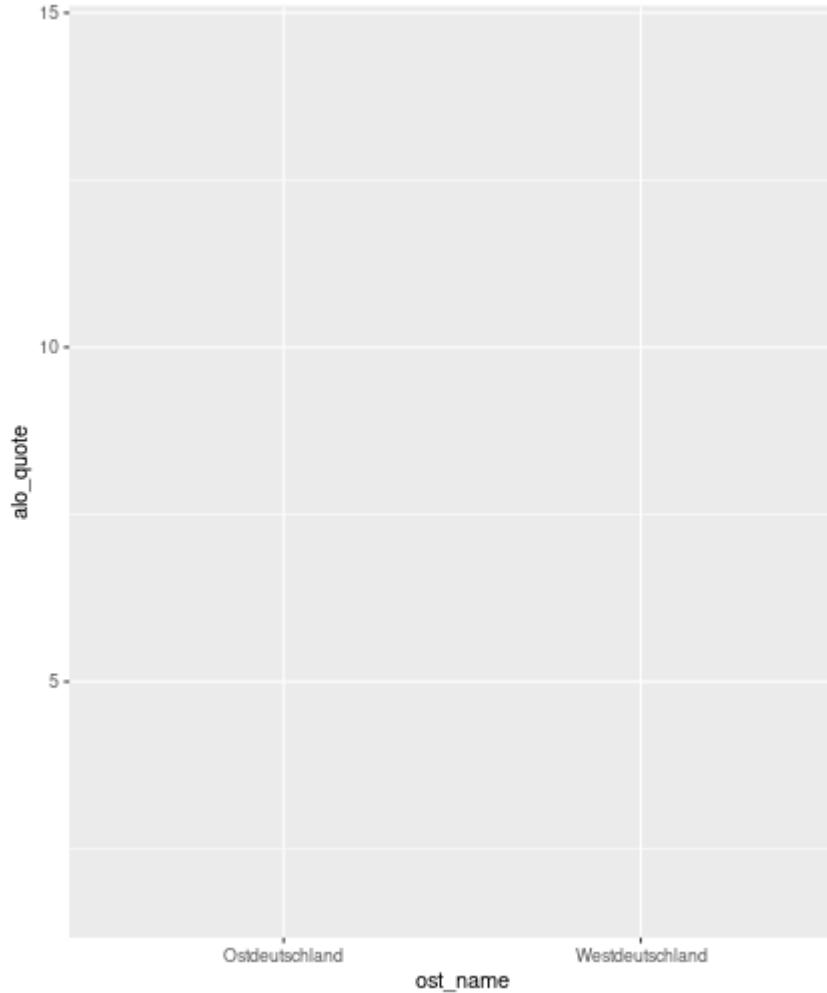
```
## # A tibble: 400 x 4
##   alo_quote landkreis_name  bundesland_name  ost_name
##       <dbl> <chr>          <chr>          <chr>
## 1 6.75   Flensburg      Schleswig-Holstein Westdeutschland
## 2 5.93   Kiel           Schleswig-Holstein Westdeutschland
## 3 6.56   Lübeck         Schleswig-Holstein Westdeutschland
## 4 6.46   Neumünster    Schleswig-Holstein Westdeutschland
## 5 6.24   Dithmarschen   Schleswig-Holstein Westdeutschland
## 6 7.20   Herzogtum Lauenburg Schleswig-Holstein Westdeutschland
## 7 4.78   Nordfriesland  Schleswig-Holstein Westdeutschland
## 8 5.44   Ostholstein    Schleswig-Holstein Westdeutschland
## 9 6.60   Pinneberg      Schleswig-Holstein Westdeutschland
## 10 5.93  Plön           Schleswig-Holstein Westdeutschland
## # ... with 390 more rows
```

```
gesamtdaten %>%
```

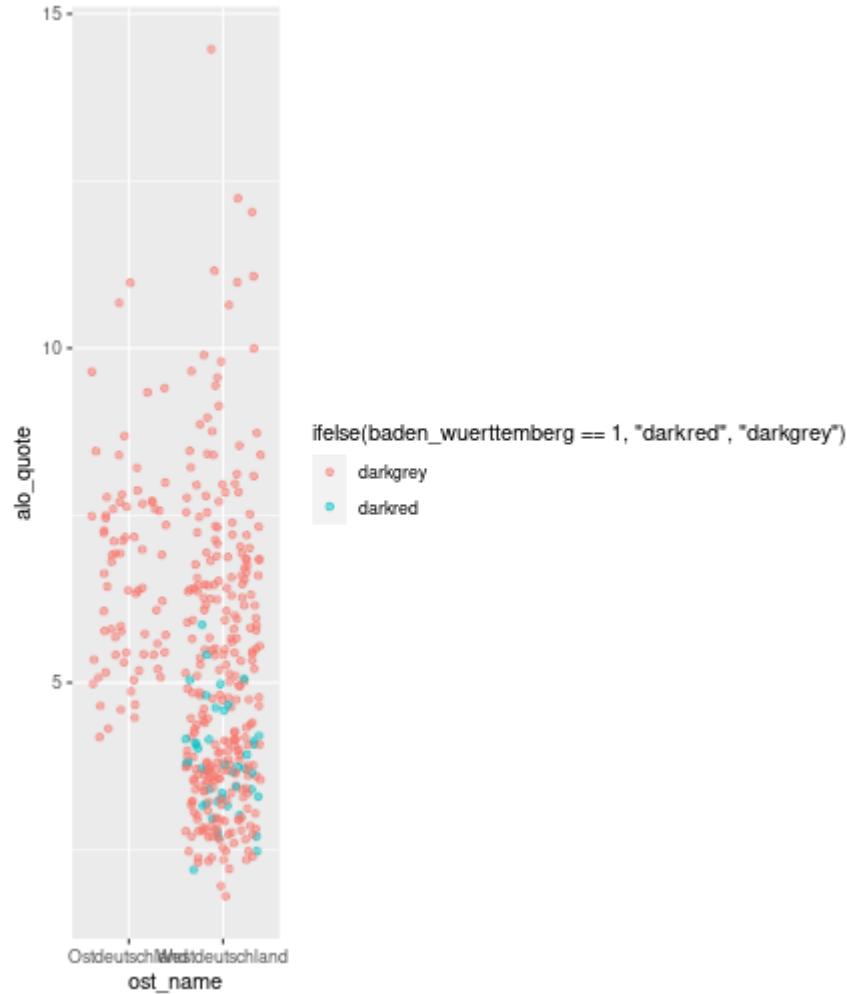
```
  select(alo_quote, landkreis_name, bundesland_name,  
         mutate(baden_wuerttemberg = as.factor(ifelse(bunde
```

```
## # A tibble: 400 x 5  
##   alo_quote landkreis_name   bundesland_name   ost_name baden_wuerttembe  
##       <dbl> <chr>           <chr>           <chr>   <fct>  
## 1 6.75 Flensburg   Schleswig-Holstein Westdeut... 0  
## 2 5.93 Kiel        Schleswig-Holstein Westdeut... 0  
## 3 6.56 Lübeck     Schleswig-Holstein Westdeut... 0  
## 4 6.46 Neumünster Schleswig-Holstein Westdeut... 0  
## 5 6.24 Dithmarschen Schleswig-Holstein Westdeut... 0  
## 6 7.20 Herzogtum Lauenburg Schleswig-Holstein Westdeut... 0  
## 7 4.78 Nordfriesland Schleswig-Holstein Westdeut... 0  
## 8 5.44 Ostholstein Schleswig-Holstein Westdeut... 0  
## 9 6.60 Pinneberg   Schleswig-Holstein Westdeut... 0  
## 10 5.93 Plön       Schleswig-Holstein Westdeut... 0  
## # i 390 more rows
```

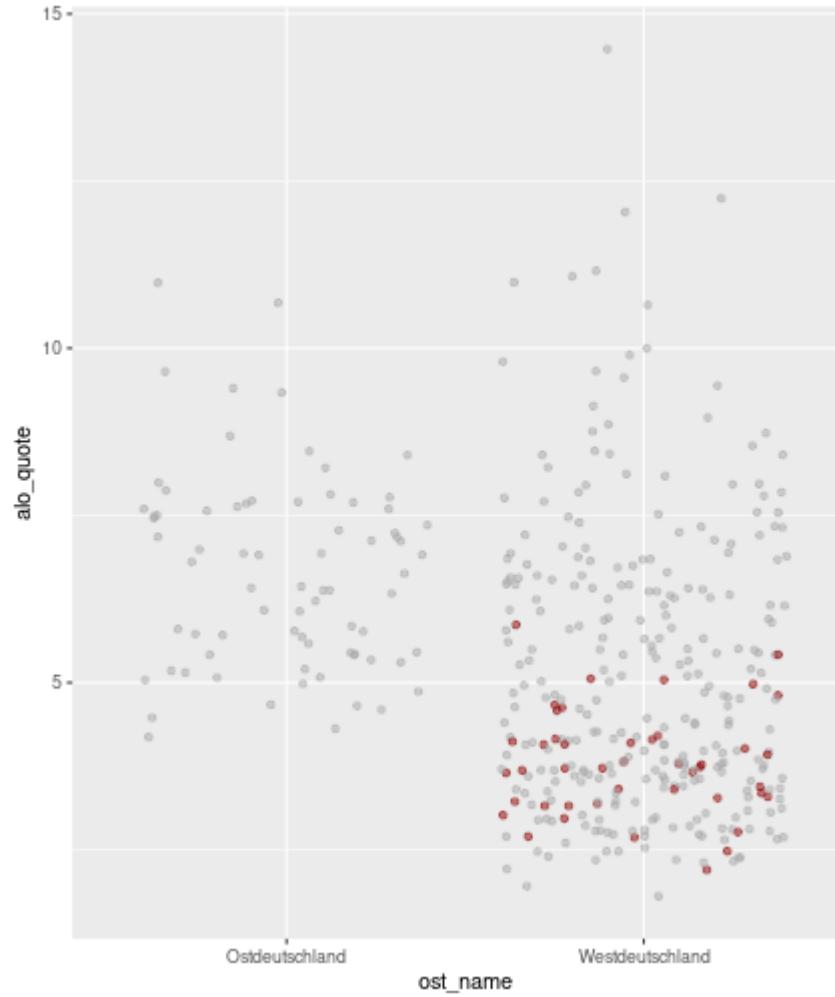
```
gesamtdataen %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
  ggplot(aes(x = ost_name, y=alo_quote))
```



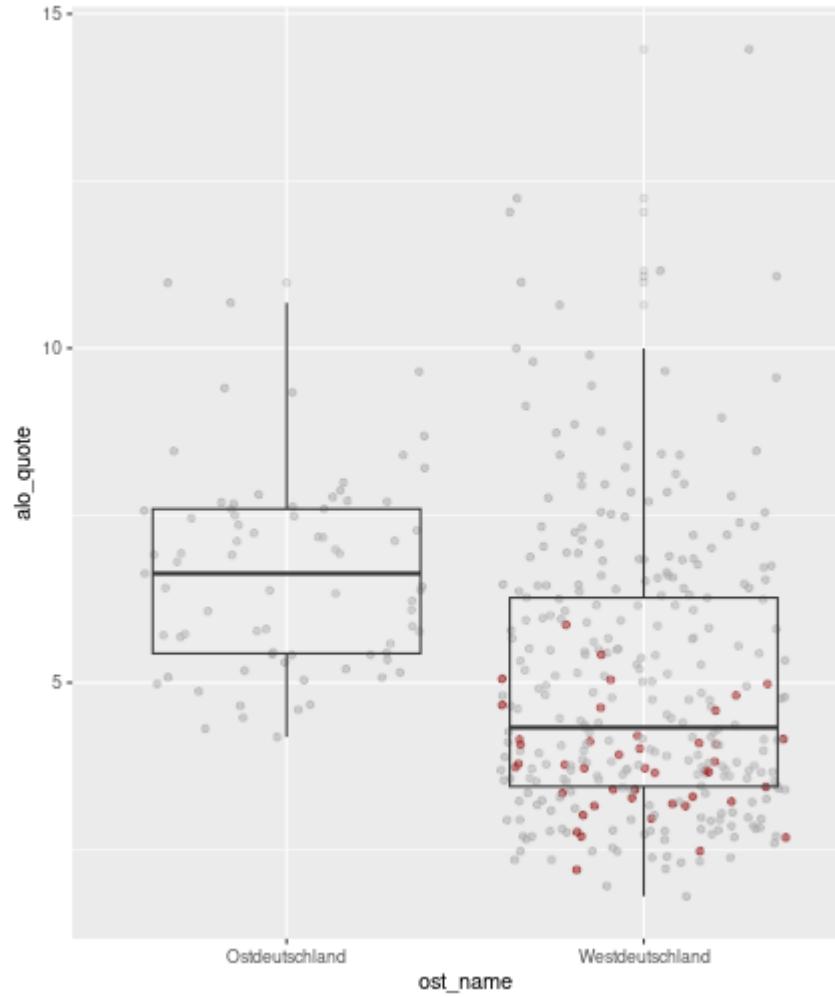
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
```



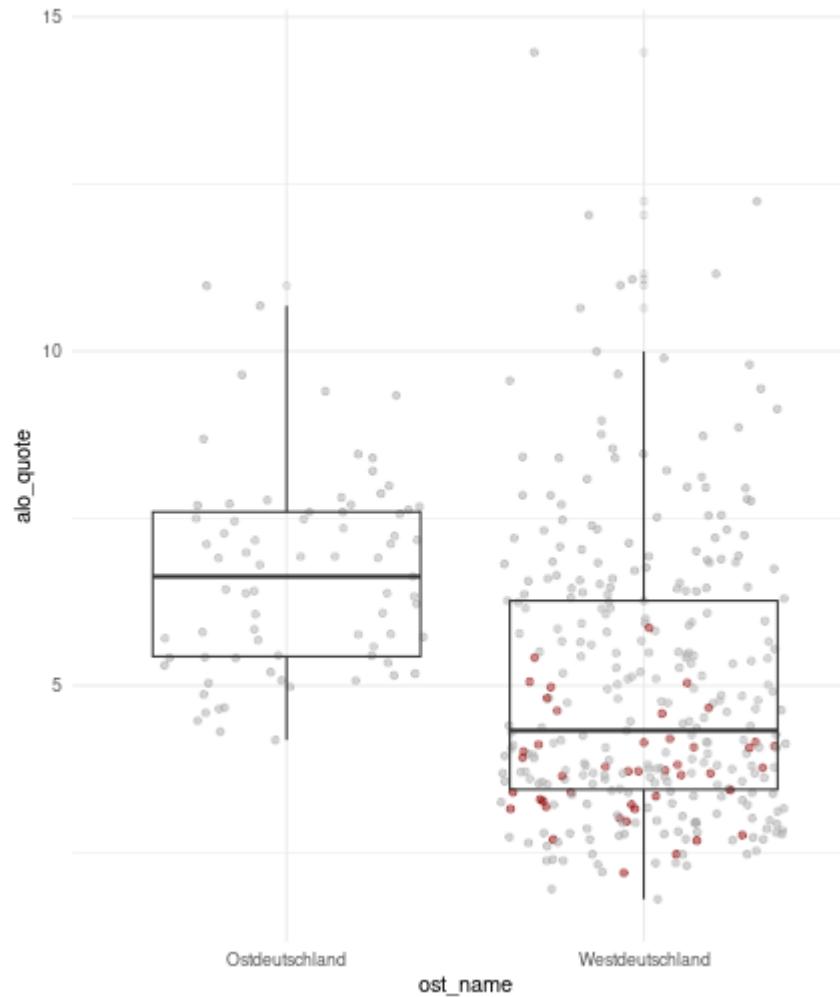
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity()
```



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



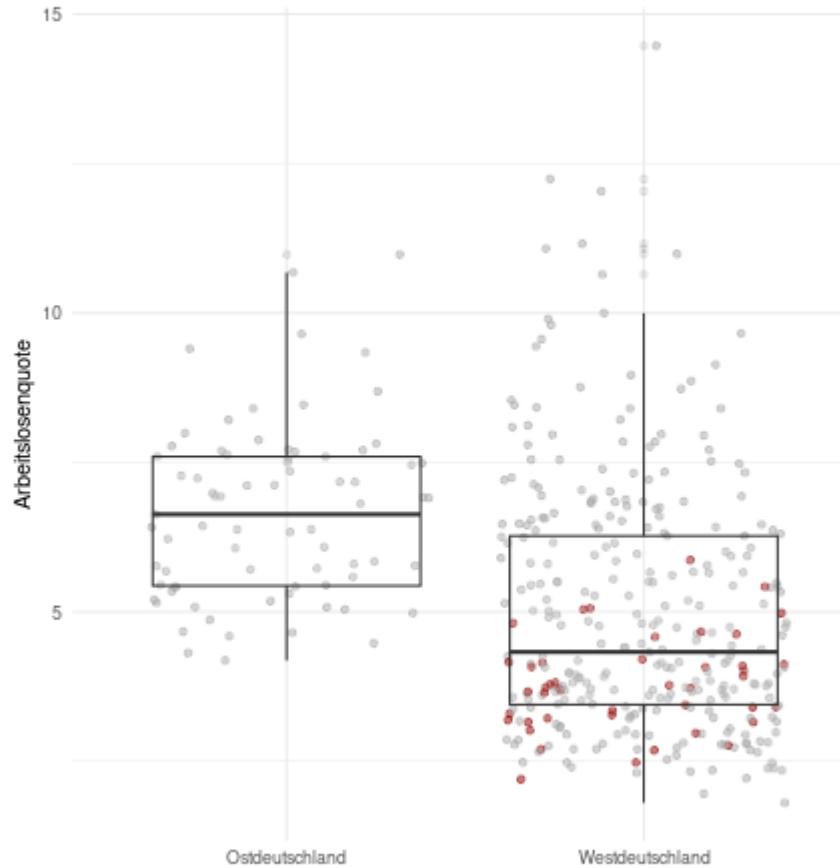
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity() +
  geom_boxplot(alpha = 0.1) +
  theme_minimal()
```



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

## Arbeitslosenquote in Deutschland

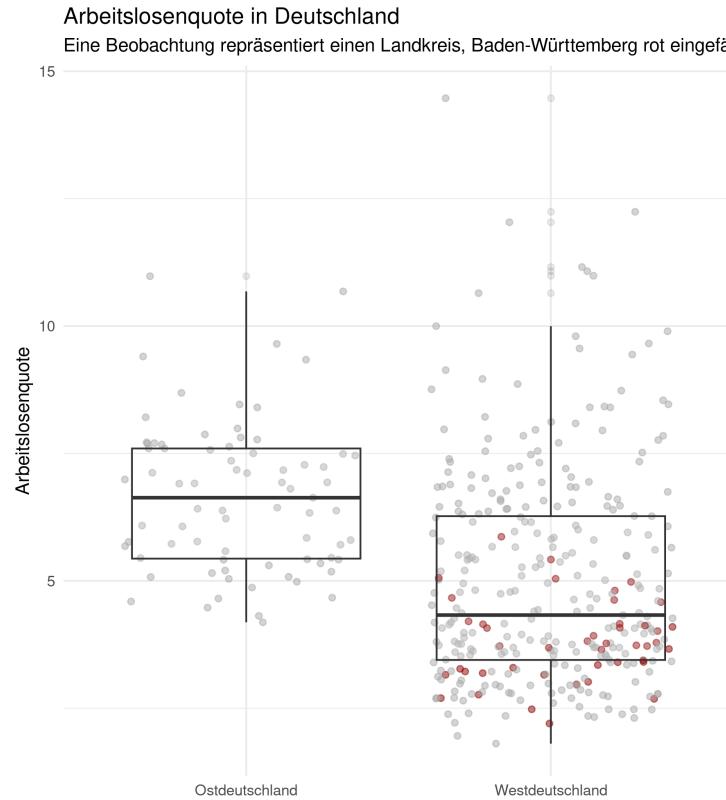
Eine Beobachtung repräsentiert einen Landkreis, Baden-Württemberg rot eingefärbt



Quelle: Daten der Agentur für Arbeit aus dem Jahr 2021

# Arbeitslosenquote

## Beschreiben Sie das gezeigte Schaubild



Quelle: Daten der Agentur für Arbeit aus dem Jahr 2021

# 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: 11,542 × 8
##   Regionalschluessel Jahr      bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>      <dbl>      <dbl>      <dbl> <chr>
## 1 08111           1992 32946884000  593628  55501. Baden-Württembe
## 2 08115           1992 12090930000  343190  35231. Baden-Württembe
## 3 08116           1992 12275605000  487370  25187. Baden-Württembe
## 4 08117           1992 5062037000  248688  20355. Baden-Württembe
## 5 08118           1992 11714160000  475248  24649. Baden-Württembe
## 6 08119           1992 8500405000  389670  21814. Baden-Württembe
## 7 08121           1992 4219259000  118566  35586. Baden-Württembe
## 8 08125           1992 6073525000  283163  21449. Baden-Württembe
## 9 08126           1992 2273334000  96072   23663. Baden-Württembe
## 10 08127          1992 3432175000  169617  20235. Baden-Württembe
## # i 11,532 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

```
## # A tibble: 8,756 x 8
##   Regionalschlüssel Jahr      bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>     <dbl>     <dbl>     <dbl> <chr>
## 1 08111           2000 35273886000  571528  61719. Baden-Württembe
## 2 08115           2000 13867882000  359476  38578. Baden-Württembe
## 3 08116           2000 14404617000  492914  29223. Baden-Württembe
## 4 08117           2000 6000420000  253970  23626. Baden-Württembe
## 5 08118           2000 14657540000  492014  29791. Baden-Württembe
## 6 08119           2000 10367512000  403830  25673. Baden-Württembe
## 7 08121           2000 5273634000  115590  45624. Baden-Württembe
## 8 08125           2000 8453750000  316406  26718. Baden-Württembe
## 9 08126           2000 3083030000  106494  28950. Baden-Württembe
## 10 08127          2000 4503834000  184222  24448. Baden-Württembe
## # i 8,746 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

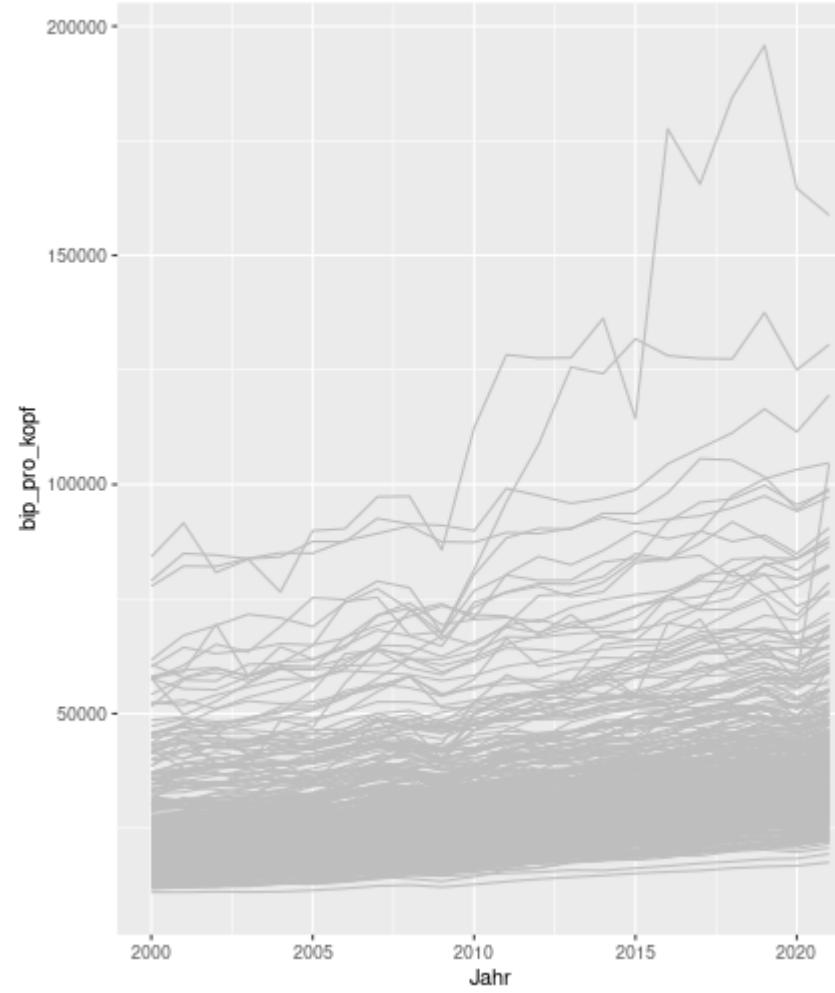
```
## # A tibble: 8,756 x 8
## # Groups:   ost_name, Jahr [44]
##   Regionalschluessel Jahr      bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl>    <chr>
## 1 08111           2000 35273886000  571528  61719. Baden-Württembe
## 2 08115           2000 13867882000  359476  38578. Baden-Württembe
## 3 08116           2000 14404617000  492914  29223. Baden-Württembe
## 4 08117           2000 6000420000  253970  23626. Baden-Württembe
## 5 08118           2000 14657540000  492014  29791. Baden-Württembe
## 6 08119           2000 10367512000  403830  25673. Baden-Württembe
## 7 08121           2000 5273634000  115590  45624. Baden-Württembe
## 8 08125           2000 8453750000  316406  26718. Baden-Württembe
## 9 08126           2000 3083030000  106494  28950. Baden-Württembe
## 10 08127          2000 4503834000  184222  24448. Baden-Württembe
## # i 8,746 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

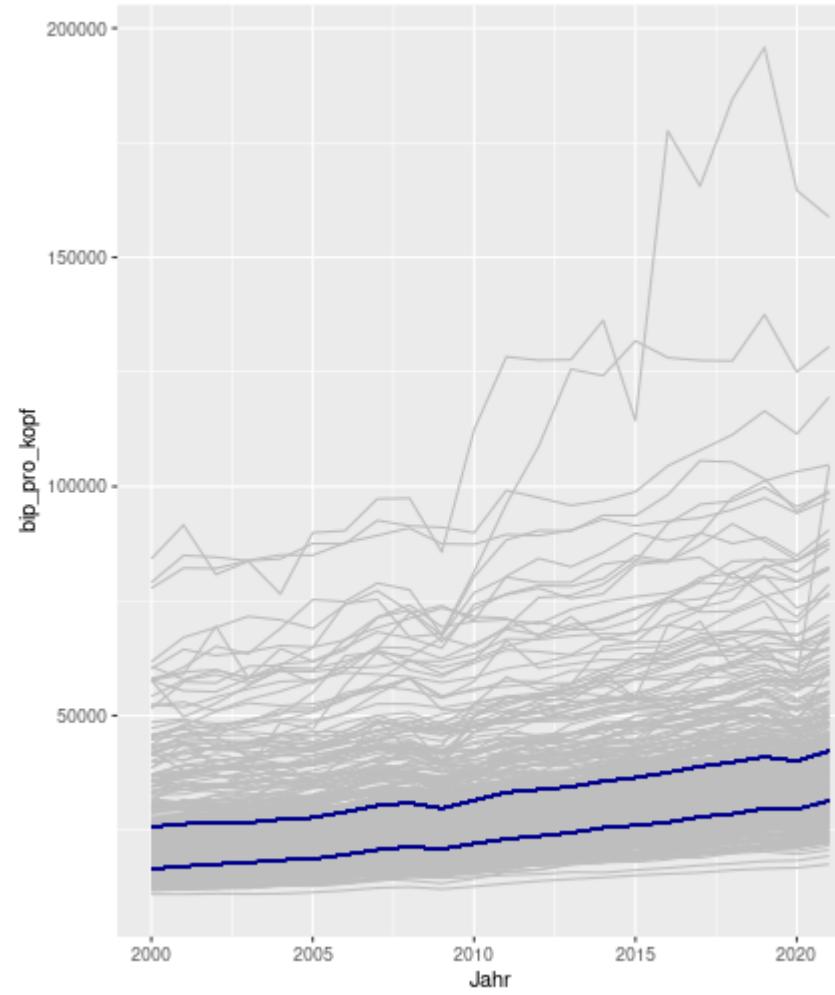
```
## # A tibble: 8,756 x 10
## # Groups:   ost_name, Jahr [44]
##   Regionalschluessel Jahr      bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl>    <chr>
## 1 08111           2000 35273886000  571528  61719. Baden-Württembe
## 2 08115           2000 13867882000  359476  38578. Baden-Württembe
## 3 08116           2000 14404617000  492914  29223. Baden-Württembe
## 4 08117           2000 6000420000  253970  23626. Baden-Württembe
## 5 08118           2000 14657540000  492014  29791. Baden-Württembe
## 6 08119           2000 10367512000  403830  25673. Baden-Württembe
## 7 08121           2000 5273634000  115590  45624. Baden-Württembe
## 8 08125           2000 8453750000  316406  26718. Baden-Württembe
## 9 08126           2000 3083030000  106494  28950. Baden-Württembe
## 10 08127          2000 4503834000  184222  24448. Baden-Württembe
## # i 8,746 more rows
## # i 4 more variables: landkreis_name <chr>, ost_name <chr>, durchschnitt <dbl>
## #   ulm <dbl>
```

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

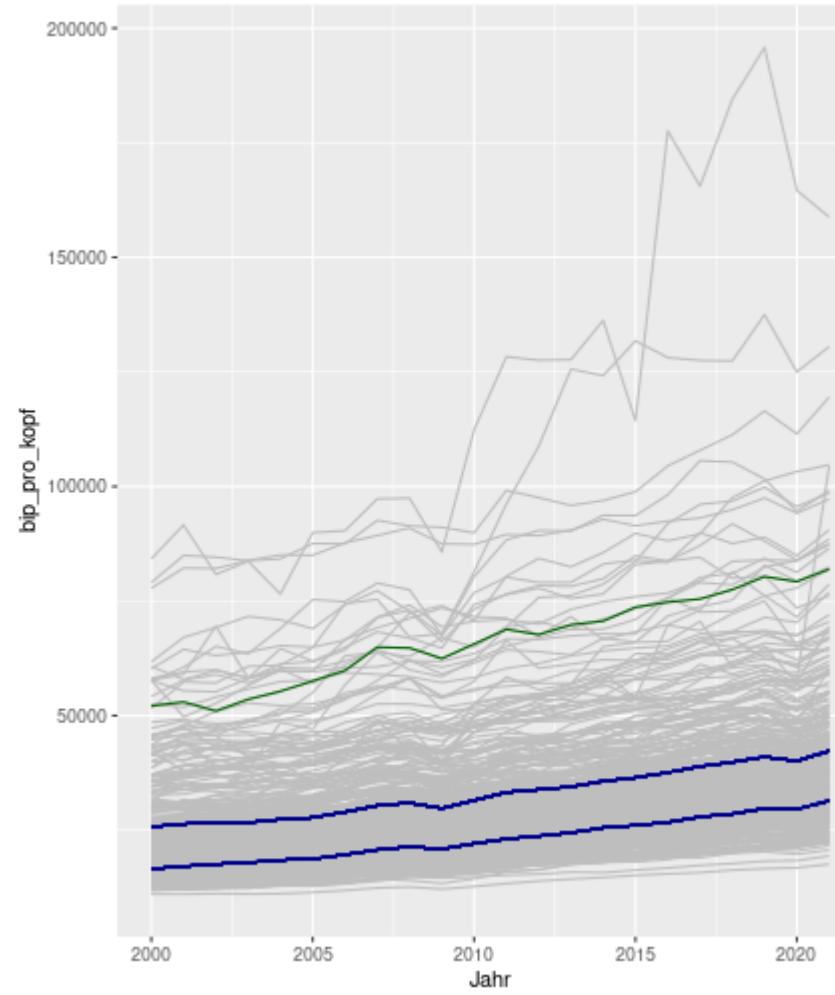
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_-
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
```



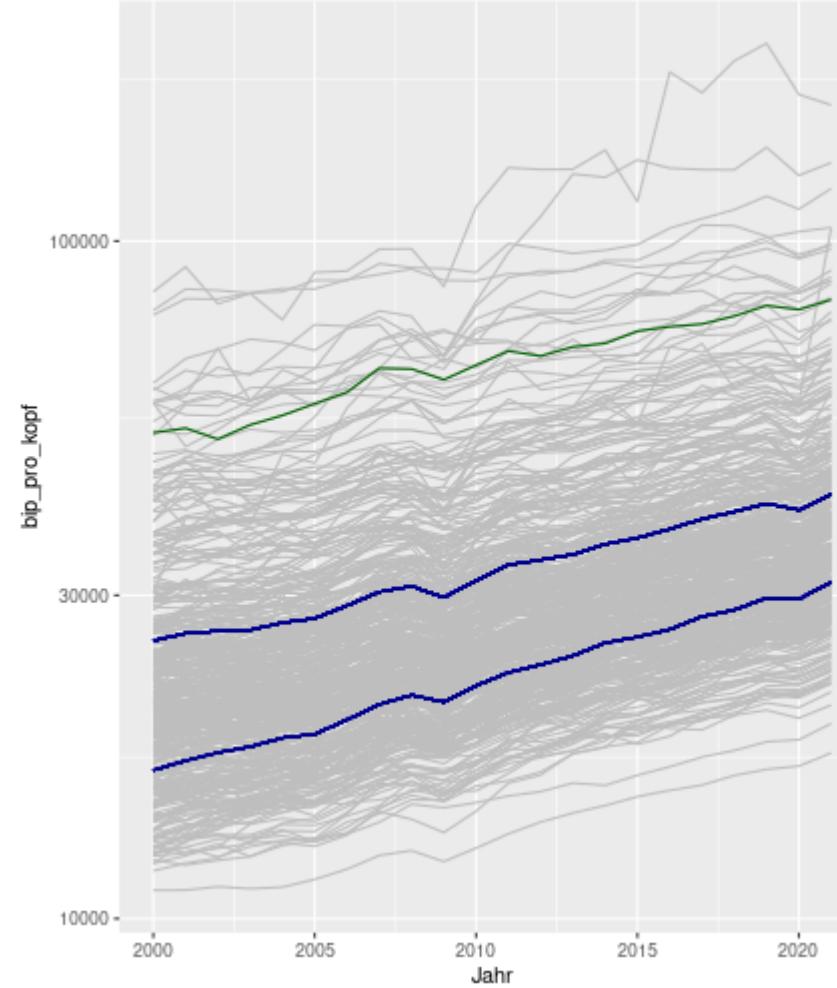
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_-
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt,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



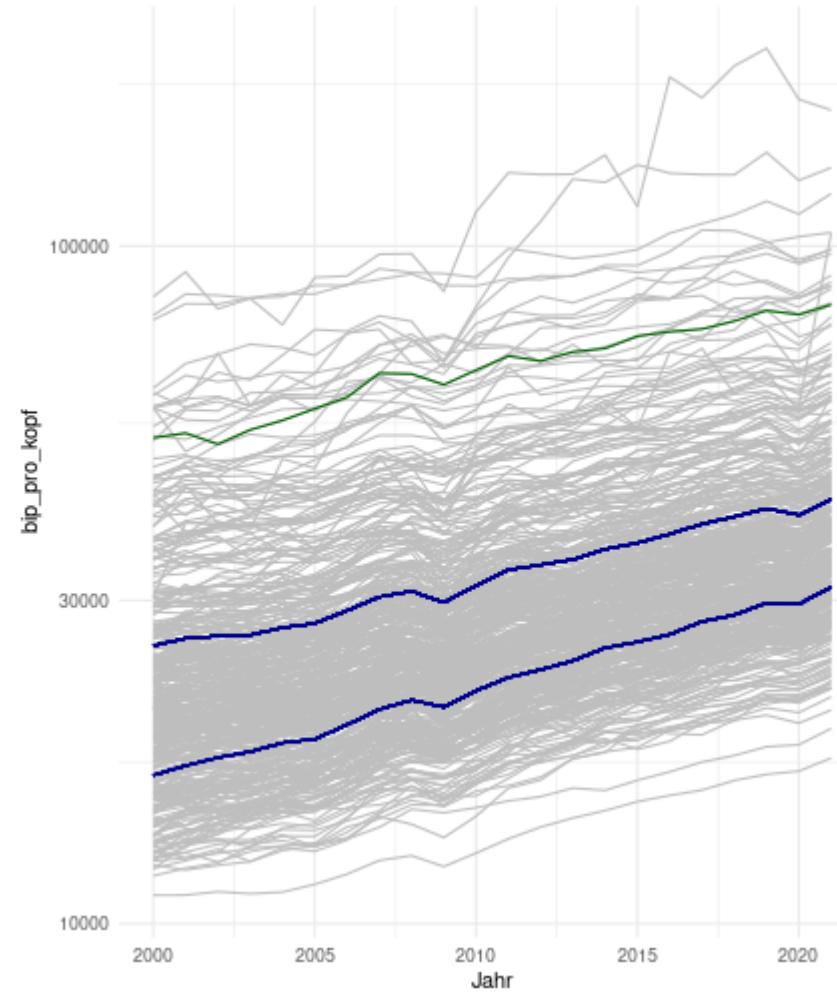
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0),
         Regionals = 0)
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group = ost_name),
            alpha = 0.1) +
  geom_line(aes(x = Jahr, y = durchschnitt, group = ost_name),
            alpha = 0.1) +
  geom_line(aes(x = Jahr, y = ulm, group = Regionals),
            alpha = 0.1) +
  geom_line(aes(x = Jahr, y = Regionals, group = Regionals),
            alpha = 0.1)
```



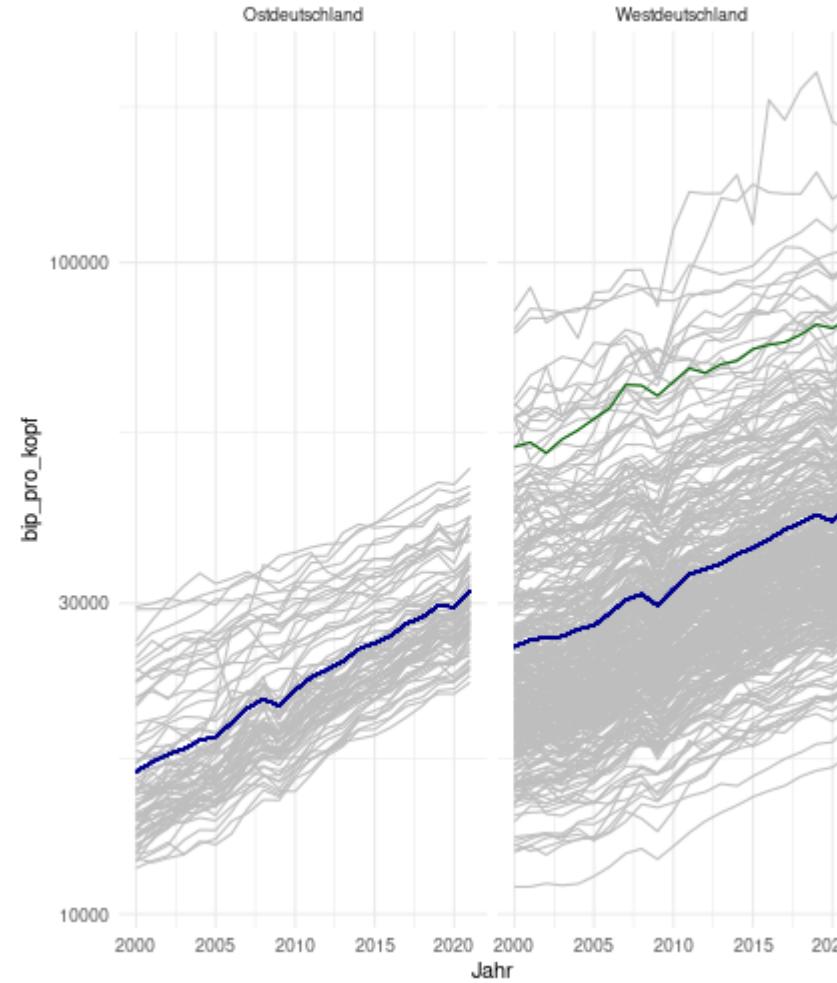
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0),
         Regionals = 0)
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group = ost_name),
            color = "#AAAAAA", size = 0.5) +
  geom_line(aes(x = Jahr, y = durchschnitt, group = 1),
            color = "#0070C0", size = 1) +
  geom_line(aes(x = Jahr, y = ulm, group = Regionals),
            color = "#00A050", size = 1) +
  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, 0),
         Regionals = 0)
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group = ost_name),
            size = 1) +
  geom_line(aes(x = Jahr, y = durchschnitt, group = 1),
            size = 1) +
  geom_line(aes(x = Jahr, y = ulm, group = Regionals),
            size = 1) +
  scale_y_continuous(trans = "log10") +
  theme_minimal()
```



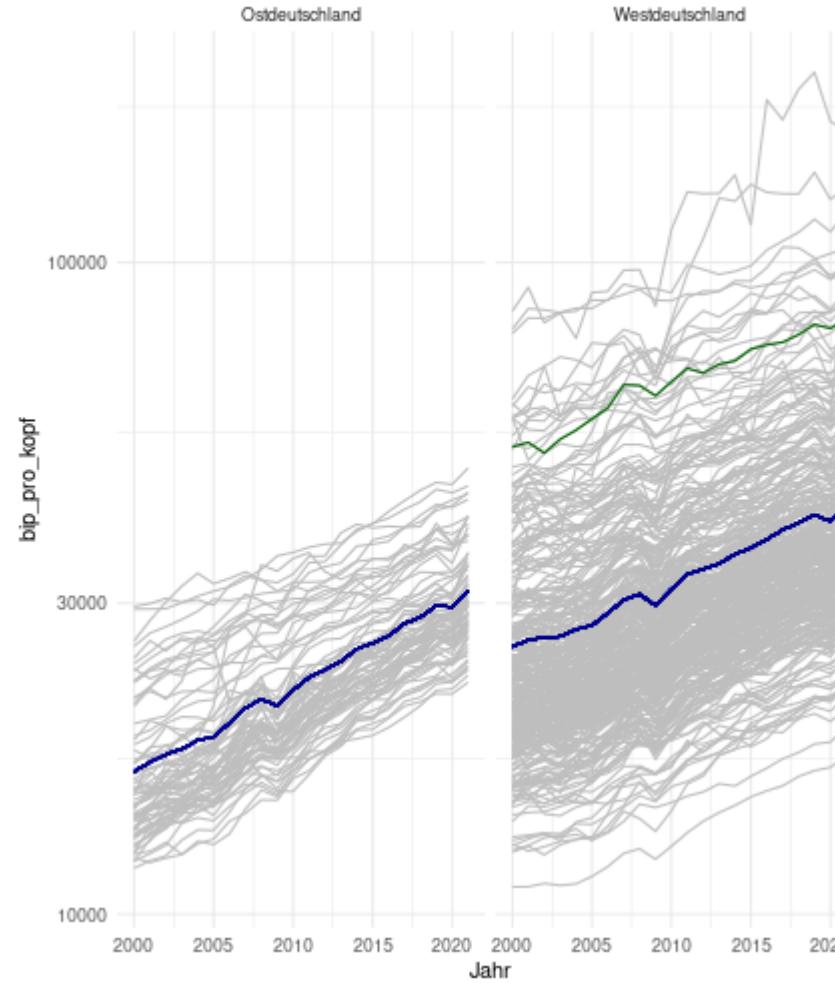
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0),
         Regionals = 0)
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group = ost_name),
            size = 1) +
  geom_line(aes(x = Jahr, y = durchschnitt, group = ost_name),
            size = 1) +
  geom_line(aes(x = Jahr, y = ulm, group = Regionals),
            size = 1) +
  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, 0),
         Regionals = 0)
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group = ost_name),
            size = 1) +
  geom_line(aes(x = Jahr, y = durchschnitt, group = ost_name),
            size = 1) +
  geom_line(aes(x = Jahr, y = ulm, group = Regionals),
            size = 1) +
  scale_y_continuous(trans = "log10") +
  theme_minimal() +
  facet_wrap(ost_name ~ .) +
  theme(legend.position = "none")

```

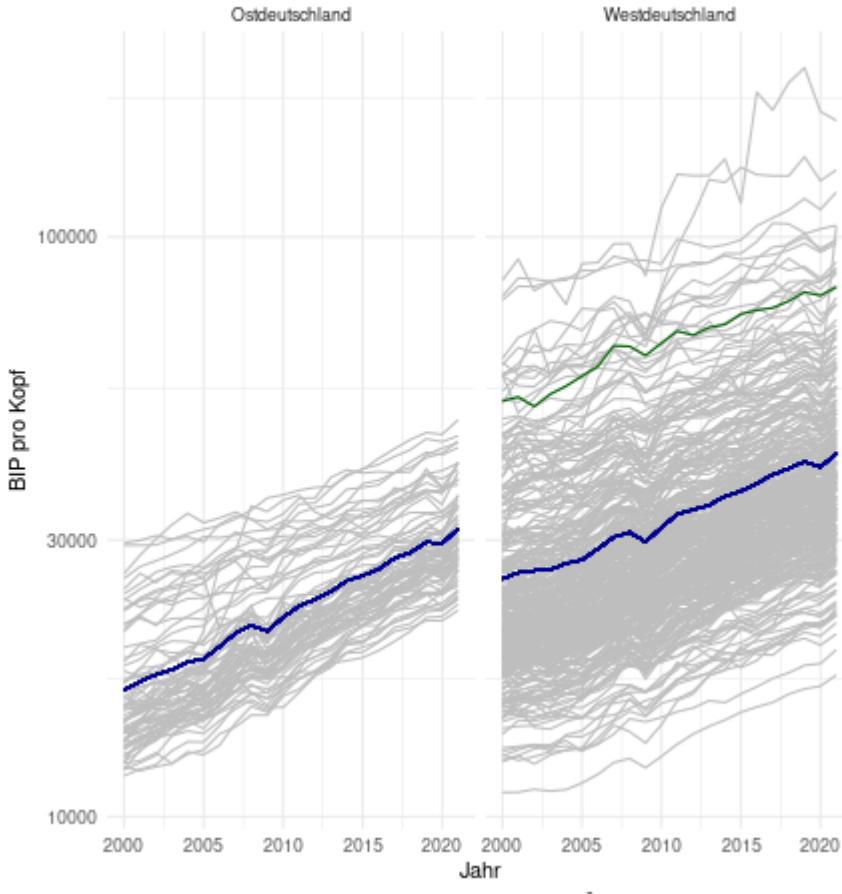


```

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),
         ost_name = ifelse(ost_name == "Ostdeutschland", "Ostdeutschland", ost_name),
         ost_name = ifelse(ost_name == "Westdeutschland", "Westdeutschland", ost_name))
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group = ost_name),
            color = "#333399", size = 1) +
  geom_line(aes(x = Jahr, y = durchschnitt, group = ost_name),
            color = "#333399", size = 1) +
  geom_line(aes(x = Jahr, y = ulm, group = ost_name),
            color = "#339933", size = 1) +
  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 Landen",
       subtitle = "Durchschnittswerte in Dunkelblau, Ulm in Dunkelgrün",
       caption = "Quelle: Daten der Statistischen Ämter der Länder und des Bundes",
       x = "Jahr",
       y = "BIP pro Kopf")

```

Ein Vergleich des BIP pro Kopf von ost- und westdeutschen Landen  
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 31384€ den Wert, welchen die westdeutschen Landkreise durchschnittlich in 2008 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 31384€ den Wert, welchen die westdeutschen Landkreise durchschnittlich in 2008 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:

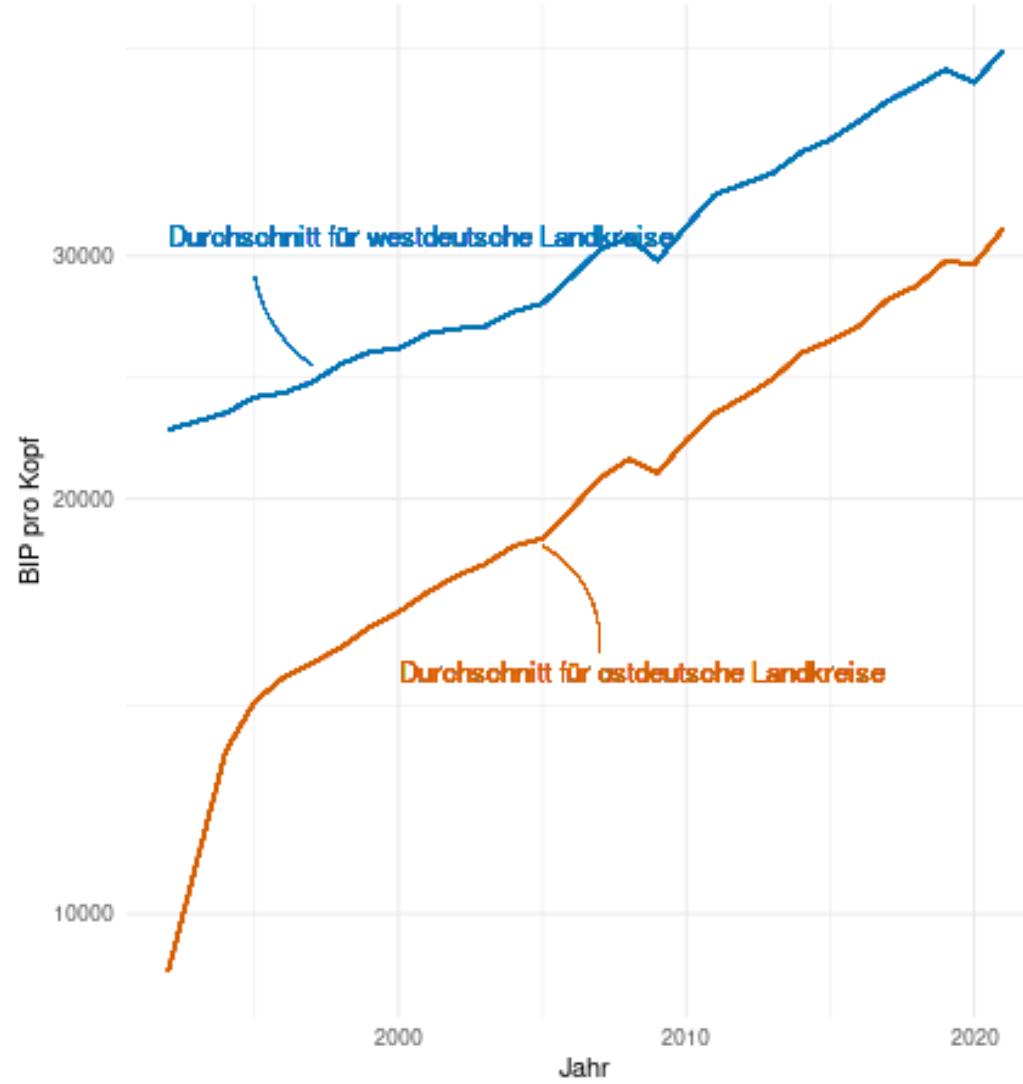
- ✚ Eine Wachstumsprozess im BIP pro Kopf findet in allen Landkreisen statt, jedoch gibt es für die ostdeutschen Landkreise, welche deutlich niedriger gestartet sind, keinen erkennbaren Anpassungsprozess in Form eines schnelleren Wachstums
- ✚ Wir sehen auch keinen Anpassungsprozess der Landkreise in 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 2021



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 2021

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 2021

berechnen und visualisieren.

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

```
bip_zeitreihe_namen
```

```
## # A tibble: 11,542 × 8
##   Regionalschluessel Jahr      bip_einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1 08111           1992 32946884000  593628  55501. Baden-Württembe
## 2 08115           1992 12090930000  343190  35231. Baden-Württembe
## 3 08116           1992 12275605000  487370  25187. Baden-Württembe
## 4 08117           1992 5062037000  248688  20355. Baden-Württembe
## 5 08118           1992 11714160000  475248  24649. Baden-Württembe
## 6 08119           1992 8500405000  389670  21814. Baden-Württembe
## 7 08121           1992 4219259000  118566  35586. Baden-Württembe
## 8 08125           1992 6073525000  283163  21449. Baden-Württembe
## 9 08126           1992 2273334000  96072   23663. Baden-Württembe
## 10 08127          1992 3432175000  169617  20235. Baden-Württembe
## # i 11,532 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

```
## # A tibble: 11,542 x 8  
## # Groups:   Regionalschluessel [398]  
##   Regionalschluessel Jahr      bip einwohner bip_pro_kopf bundesland_name  
##   <chr>          <dbl>    <dbl>      <dbl>      <dbl> <chr>  
## 1 08111          1992 32946884000  593628  55501. Baden-Württembe  
## 2 08115          1992 12090930000  343190  35231. Baden-Württembe  
## 3 08116          1992 12275605000  487370  25187. Baden-Württembe  
## 4 08117          1992 5062037000  248688  20355. Baden-Württembe  
## 5 08118          1992 11714160000  475248  24649. Baden-Württembe  
## 6 08119          1992 8500405000  389670  21814. Baden-Württembe  
## 7 08121          1992 4219259000  118566  35586. Baden-Württembe  
## 8 08125          1992 6073525000  283163  21449. Baden-Württembe  
## 9 08126          1992 2273334000  96072   23663. Baden-Württembe  
## 10 08127         1992 3432175000  169617  20235. Baden-Württembe  
## # i 11,532 more rows  
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

## # A tibble: 11,542 × 8
## # Groups:   Regionalschluessel [398]
##   Regionalschluessel Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>   <dbl>      <dbl>      <dbl> <chr>
## 1 01001            1992 2555393000  86642     29494. Schleswig-Holste
## 2 01001            1994 2624290000  86287     30414. Schleswig-Holste
## 3 01001            1995 2628469000  85506     30740. Schleswig-Holste
## 4 01001            1996 2578880000  84499     30520. Schleswig-Holste
## 5 01001            1997 2725086000  83344     32697. Schleswig-Holste
## 6 01001            1998 2872187000  82112     34979. Schleswig-Holste
## 7 01001            1999 2712126000  81276     33369. Schleswig-Holste
## 8 01001            2000 2487282000  80758     30799. Schleswig-Holste
## 9 01001            2001 2465393000  80489     30630. Schleswig-Holste
## 10 01001           2002 2635779000  80414     32778. Schleswig-Holste
## # i 11,532 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

## # A tibble: 11,542 × 9
## # Groups:   Regionalschluessel [398]
##   Regionalschluessel Jahr     bip_einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>       <dbl>       <dbl>       <dbl> <chr>
## 1 01001            1992 2555393000 86642 29494. Schleswig-Holste
## 2 01001            1994 2624290000 86287 30414. Schleswig-Holste
## 3 01001            1995 2628469000 85506 30740. Schleswig-Holste
## 4 01001            1996 2578880000 84499 30520. Schleswig-Holste
## 5 01001            1997 2725086000 83344 32697. Schleswig-Holste
## 6 01001            1998 2872187000 82112 34979. Schleswig-Holste
## 7 01001            1999 2712126000 81276 33369. Schleswig-Holste
## 8 01001            2000 2487282000 80758 30799. Schleswig-Holste
## 9 01001            2001 2465393000 80489 30630. Schleswig-Holste
## 10 01001           2002 2635779000 80414 32778. Schleswig-Holste
## # i 11,532 more rows
## # i 3 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>
```

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

```
## # A tibble: 11,542 x 9
##   Regionalschluessel Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>      <dbl>      <dbl> <chr>
## 1 01001            1992 2555393000  86642  29494. Schleswig-Holste
## 2 01001            1994 2624290000  86287  30414. Schleswig-Holste
## 3 01001            1995 2628469000  85506  30740. Schleswig-Holste
## 4 01001            1996 2578880000  84499  30520. Schleswig-Holste
## 5 01001            1997 2725086000  83344  32697. Schleswig-Holste
## 6 01001            1998 2872187000  82112  34979. Schleswig-Holste
## 7 01001            1999 2712126000  81276  33369. Schleswig-Holste
## 8 01001            2000 2487282000  80758  30799. Schleswig-Holste
## 9 01001            2001 2465393000  80489  30630. Schleswig-Holste
## 10 01001           2002 2635779000  80414  32778. Schleswig-Holste
## # i 11,532 more rows
## # i 3 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>
```

```
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: 11,542 x 9
## # Groups:   ost_name, Jahr [58]
##   Regionalschluessel Jahr     bip_einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl>        <dbl>        <dbl>        <chr>
## 1 01001        1992 2555393000 86642 29494. Schleswig-Holste
## 2 01001        1994 2624290000 86287 30414. Schleswig-Holste
## 3 01001        1995 2628469000 85506 30740. Schleswig-Holste
## 4 01001        1996 2578880000 84499 30520. Schleswig-Holste
## 5 01001        1997 2725086000 83344 32697. Schleswig-Holste
## 6 01001        1998 2872187000 82112 34979. Schleswig-Holste
## 7 01001        1999 2712126000 81276 33369. Schleswig-Holste
## 8 01001        2000 2487282000 80758 30799. Schleswig-Holste
## 9 01001        2001 2465393000 80489 30630. Schleswig-Holste
## 10 01001       2002 2635779000 80414 32778. Schleswig-Holste
## # i 11,532 more rows
## # i 3 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>
```

```
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: 11,542 x 10
## # Groups:   ost_name, Jahr [58]
##   Regionalschluessel Jahr     bip_einwohner bip_pro_kopf bundesland_name
##   <chr>        <dbl>        <dbl>        <dbl>        <chr>
## 1 01001        1992 2555393000 86642 29494. Schleswig-Holste
## 2 01001        1994 2624290000 86287 30414. Schleswig-Holste
## 3 01001        1995 2628469000 85506 30740. Schleswig-Holste
## 4 01001        1996 2578880000 84499 30520. Schleswig-Holste
## 5 01001        1997 2725086000 83344 32697. Schleswig-Holste
## 6 01001        1998 2872187000 82112 34979. Schleswig-Holste
## 7 01001        1999 2712126000 81276 33369. Schleswig-Holste
## 8 01001        2000 2487282000 80758 30799. Schleswig-Holste
## 9 01001        2001 2465393000 80489 30630. Schleswig-Holste
## 10 01001       2002 2635779000 80414 32778. Schleswig-Holste
## # i 11,532 more rows
## # i 4 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>, durchschnitt <dbl>
```

```
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
```

```
bip_wachstum
```

```
## # A tibble: 11,542 x 10
##   Regionalschluessel Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>      <dbl>      <dbl> <chr>
## 1 01001            1992 2555393000  86642  29494. Schleswig-Holste
## 2 01001            1994 2624290000  86287  30414. Schleswig-Holste
## 3 01001            1995 2628469000  85506  30740. Schleswig-Holste
## 4 01001            1996 2578880000  84499  30520. Schleswig-Holste
## 5 01001            1997 2725086000  83344  32697. Schleswig-Holste
## 6 01001            1998 2872187000  82112  34979. Schleswig-Holste
## 7 01001            1999 2712126000  81276  33369. Schleswig-Holste
## 8 01001            2000 2487282000  80758  30799. Schleswig-Holste
## 9 01001            2001 2465393000  80489  30630. Schleswig-Holste
## 10 01001           2002 2635779000  80414  32778. Schleswig-Holste
## # i 11,532 more rows
## # i 4 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>, durchschnitt <dbl>
```

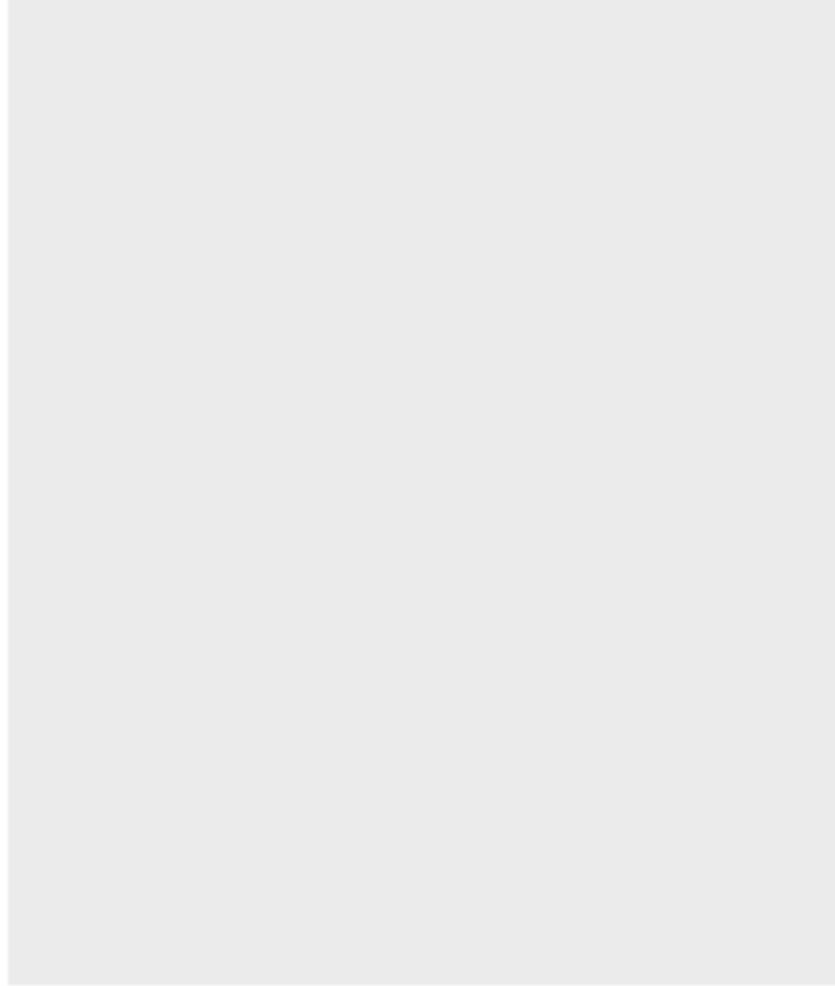
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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

```
## # A tibble: 8,756 x 10
##   Regionalschluessel Jahr     bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl> <chr>
## 1 01001            2000 2487282000 80758 30799. Schleswig-Holste
## 2 01001            2001 2465393000 80489 30630. Schleswig-Holste
## 3 01001            2002 2635779000 80414 32778. Schleswig-Holste
## 4 01001            2003 2705233000 80538 33590. Schleswig-Holste
## 5 01001            2004 2867823000 80783 35500. Schleswig-Holste
## 6 01001            2005 2900791000 80892 35860. Schleswig-Holste
## 7 01001            2006 3021890000 81052 37283. Schleswig-Holste
## 8 01001            2007 3051250000 81634 37377. Schleswig-Holste
## 9 01001            2008 3105912000 82403 37692. Schleswig-Holste
## 10 01001           2009 2993960000 82478 36300. Schleswig-Holste
## # i 8,746 more rows
## # i 4 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>, durchschnitt <dbl>
```

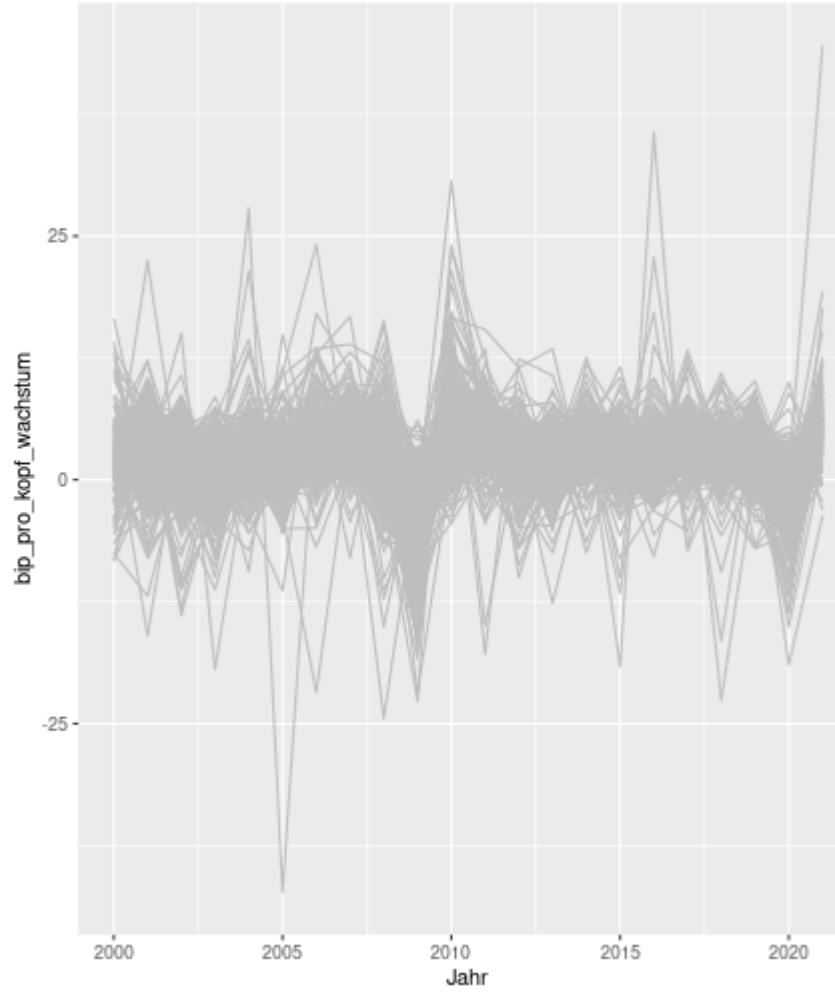
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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



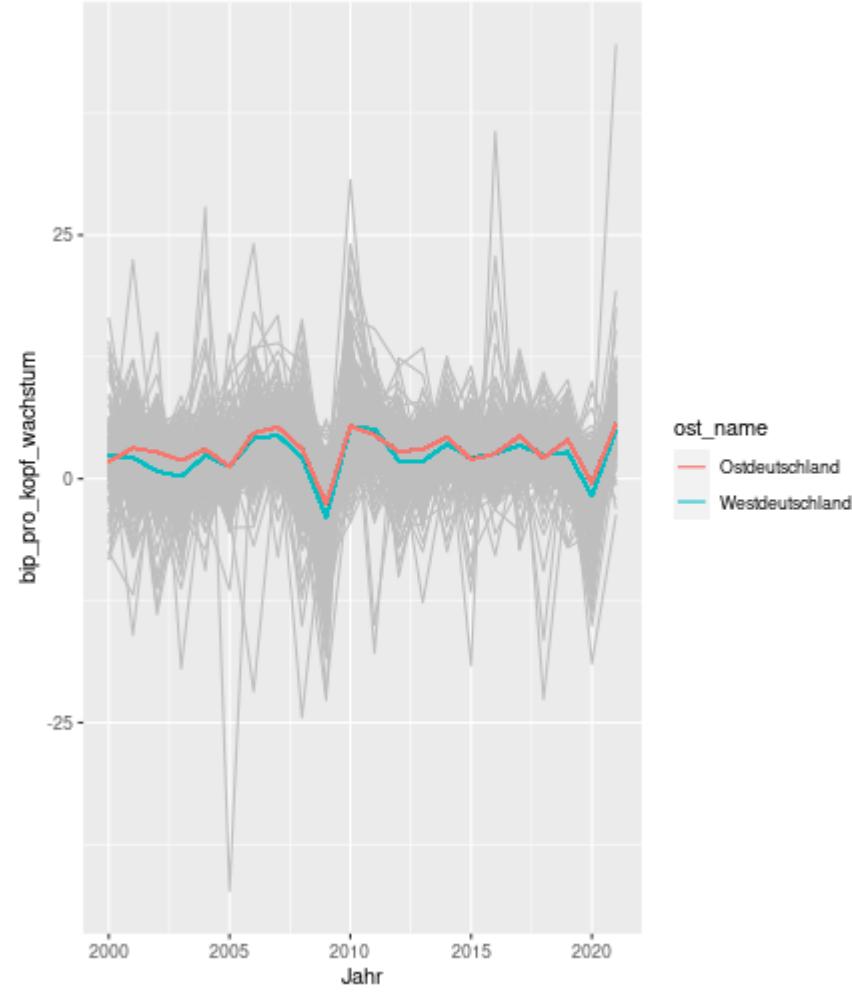
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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 =
```

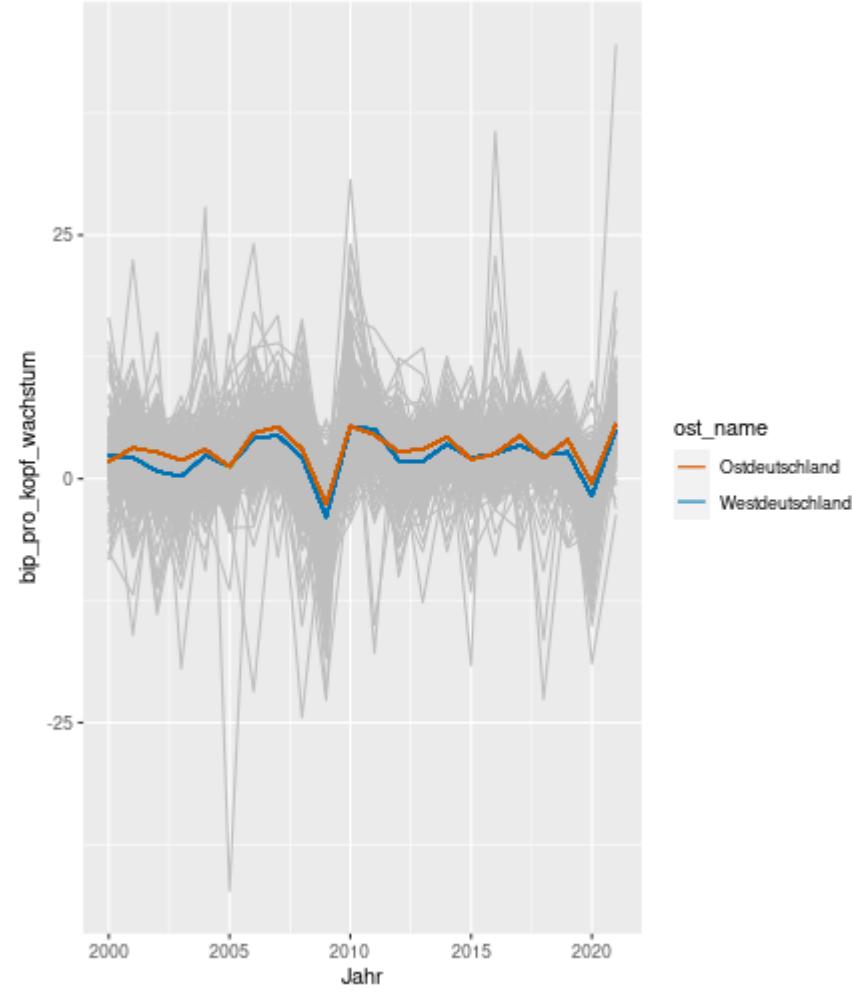


```

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"

```

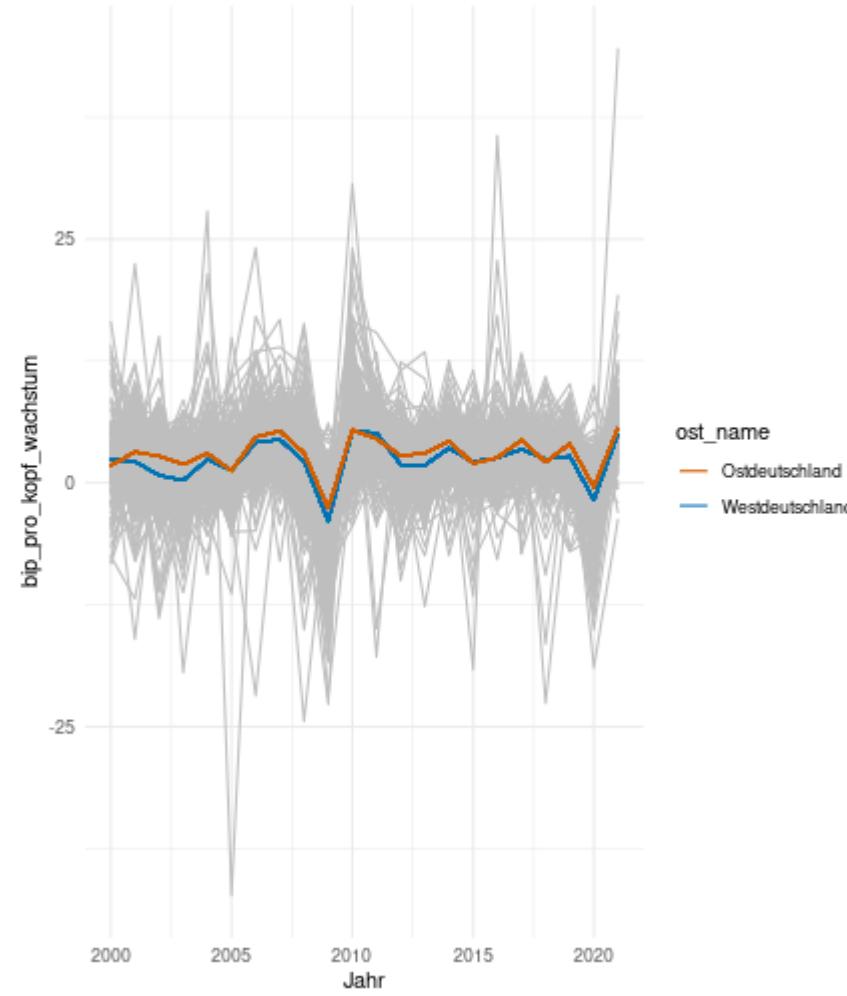


```

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()

```



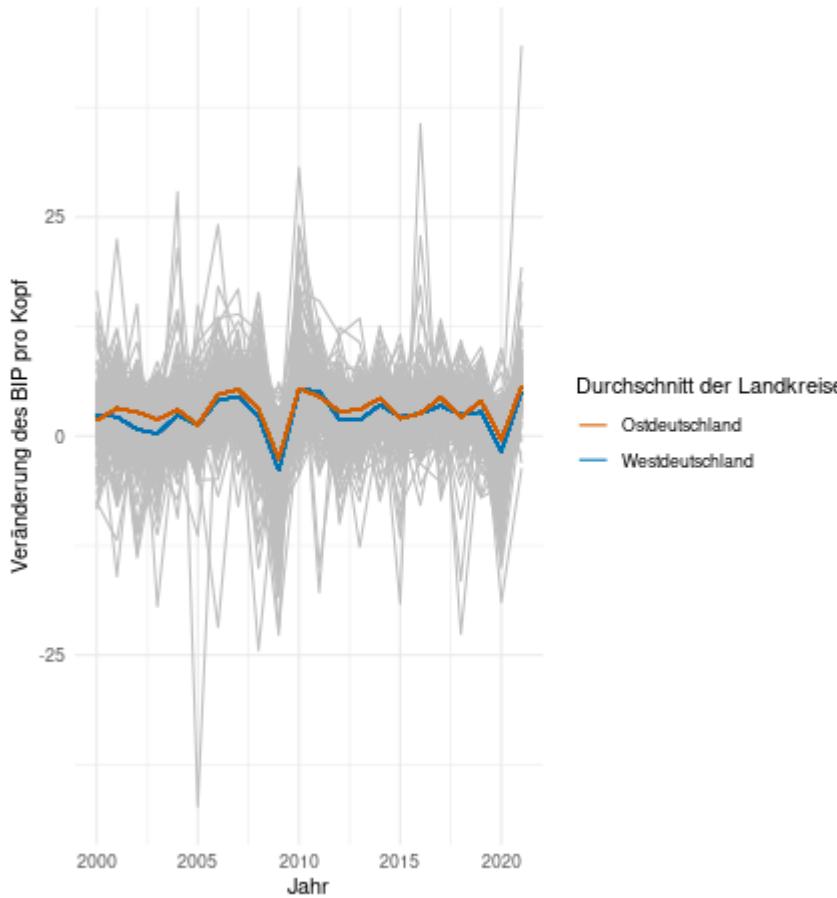
```

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



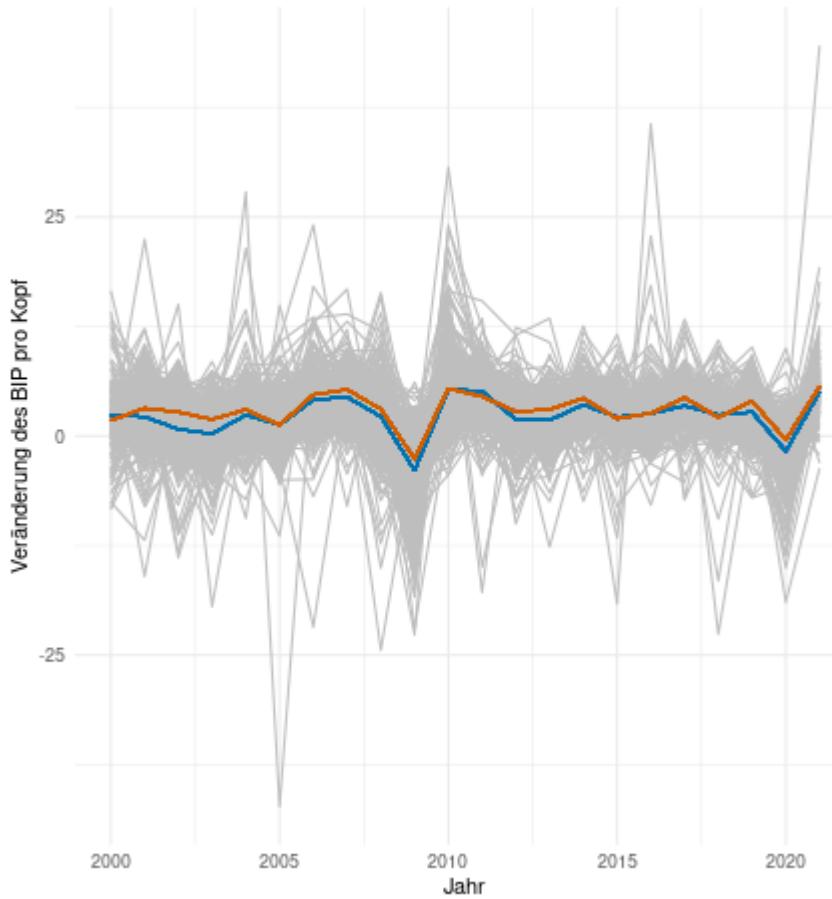
```

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 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")

```

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



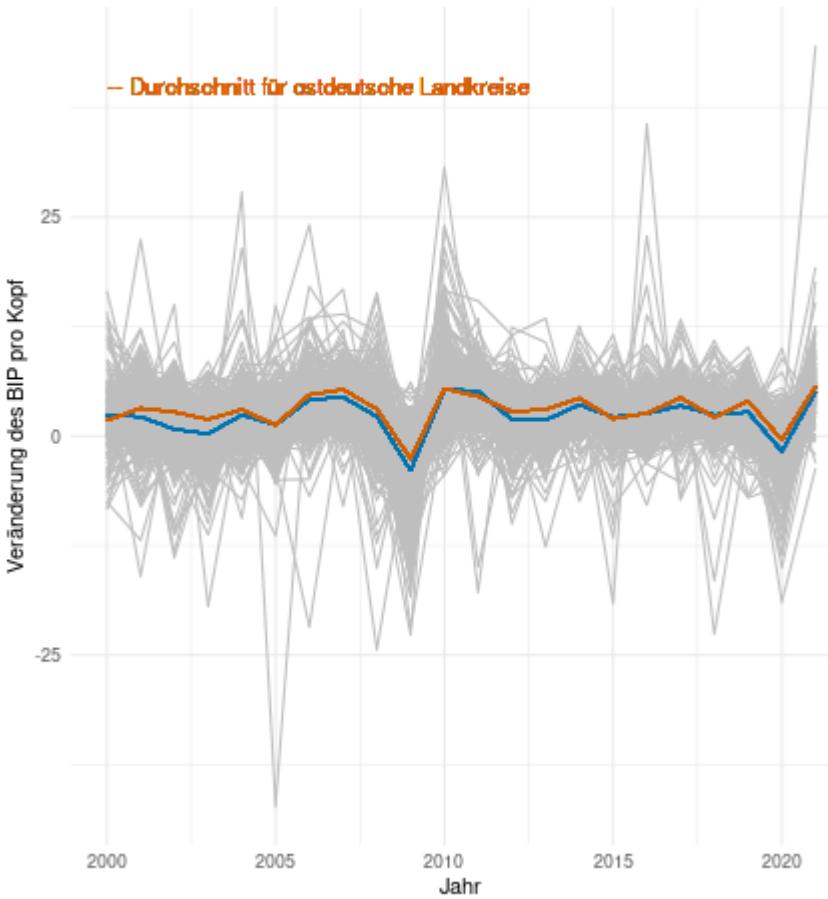
```

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 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



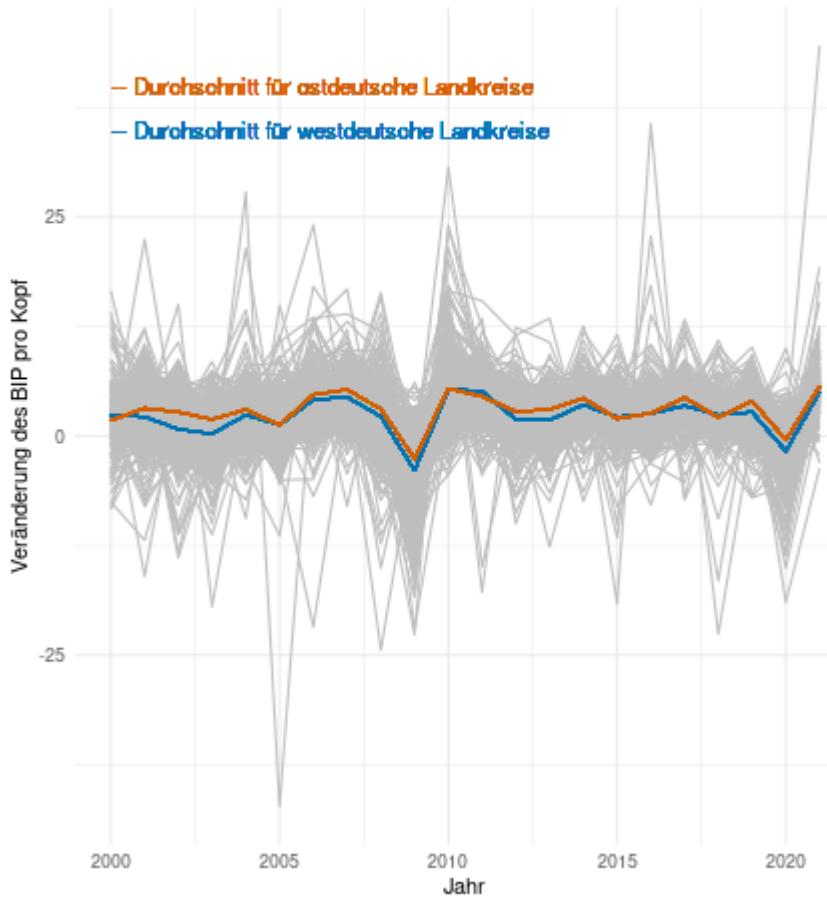
```

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: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  4369. Flensburg          01        Schleswig-Holst
## 2 01002                  11097. Kiel              01        Schleswig-Holst
## 3 01003                  9347. Lübeck            01        Schleswig-Holst
## 4 01004                  3771. Neumünster        01        Schleswig-Holst
## 5 01051                  4143. Dithmarschen        01        Schleswig-Holst
## 6 01053                  5603. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4699 Nordfriesland        01        Schleswig-Holst
## 8 01055                  5371 Ostholstein          01        Schleswig-Holst
## 9 01056                  9371. Pinneberg          01        Schleswig-Holst
## 10 01057                 2854. Plön               01        Schleswig-Holst
## # i 390 more rows
## # i 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>
```

```
gesamtdaten %>%
```

```
  group_by(ost_name)
```

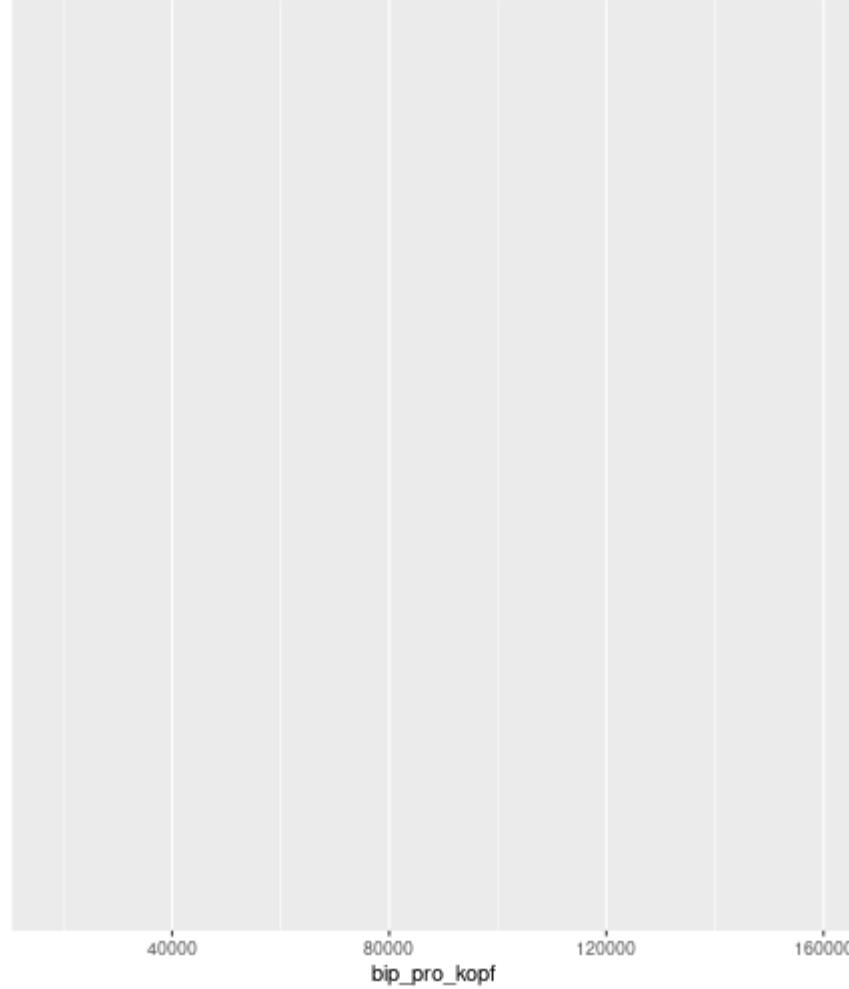
```
## # A tibble: 400 x 14
## # Groups:   ost_name [2]
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>           <chr>      <chr>
## 1 01001            4369. Flensburg            01       Schleswig-Holst
## 2 01002            11097. Kiel                01       Schleswig-Holst
## 3 01003            9347. Lübeck              01       Schleswig-Holst
## 4 01004            3771. Neumünster          01       Schleswig-Holst
## 5 01051            4143. Dithmarschen         01       Schleswig-Holst
## 6 01053            5603. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4699 Nordfriesland          01       Schleswig-Holst
## 8 01055            5371 Ostholstein           01       Schleswig-Holst
## 9 01056            9371. Pinneberg            01       Schleswig-Holst
## 10 01057           2854. Plön                 01       Schleswig-Holst
## # ... with 390 more rows
## # ... with 9 more variables: Schulden_pro_kopf_lk <dbl>, Einwohner <dbl>,
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>
```

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

```
## # A tibble: 2 × 2  
##   ost_name      durchschnitt  
##   <chr>          <dbl>  
## 1 Ostdeutschland     31384.  
## 2 Westdeutschland    42298.
```

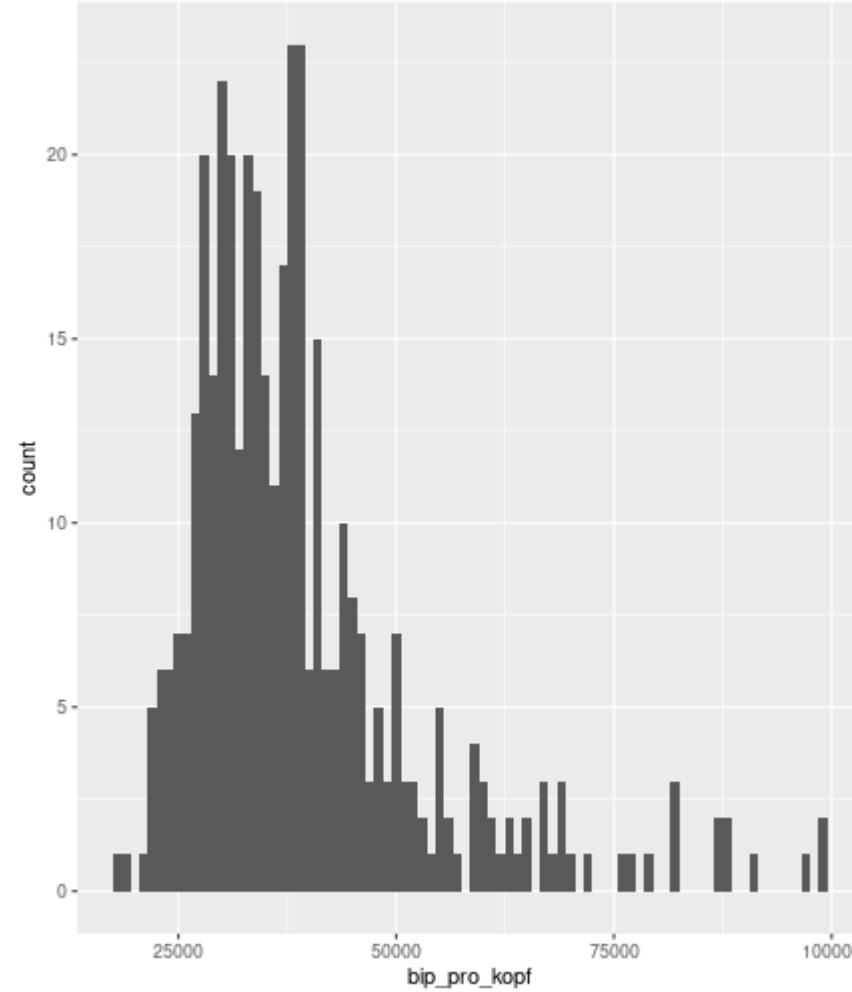
```
gesamtdataen %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel
```

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



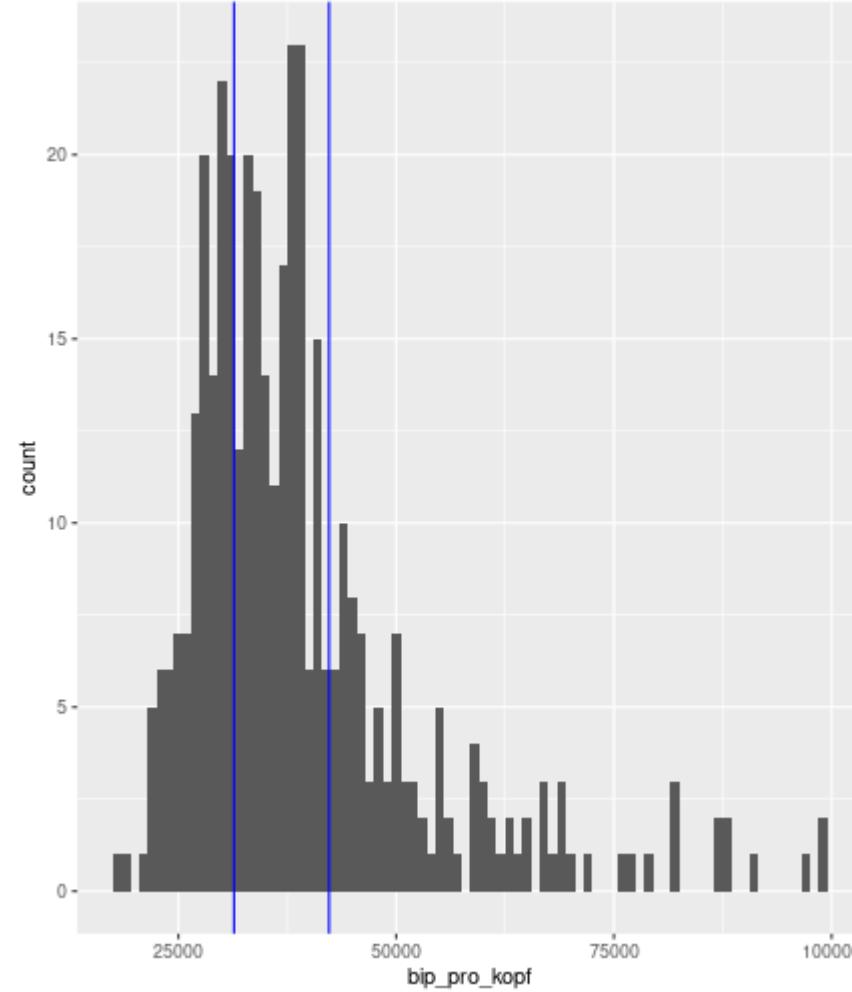
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel

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



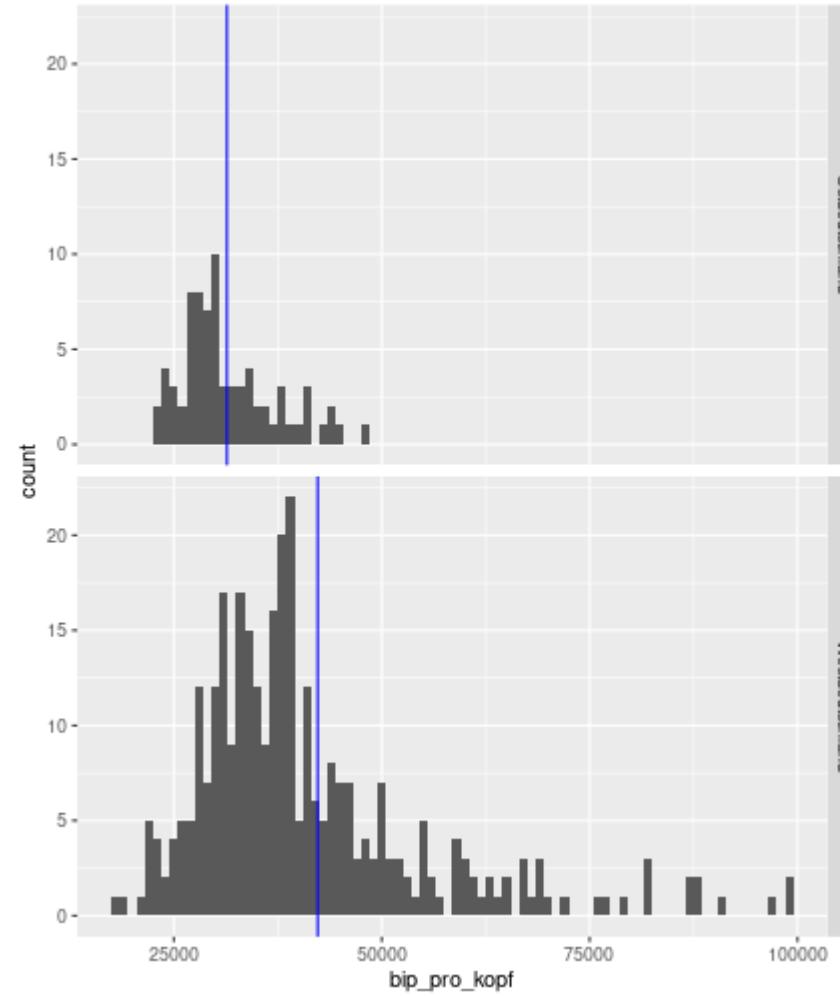
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf > 0), fill = "#333333", color = "white", binwidth = 2500) +
  geom_vline(data = mittel, aes(xintercept = durchschnitt), color = "blue")
```



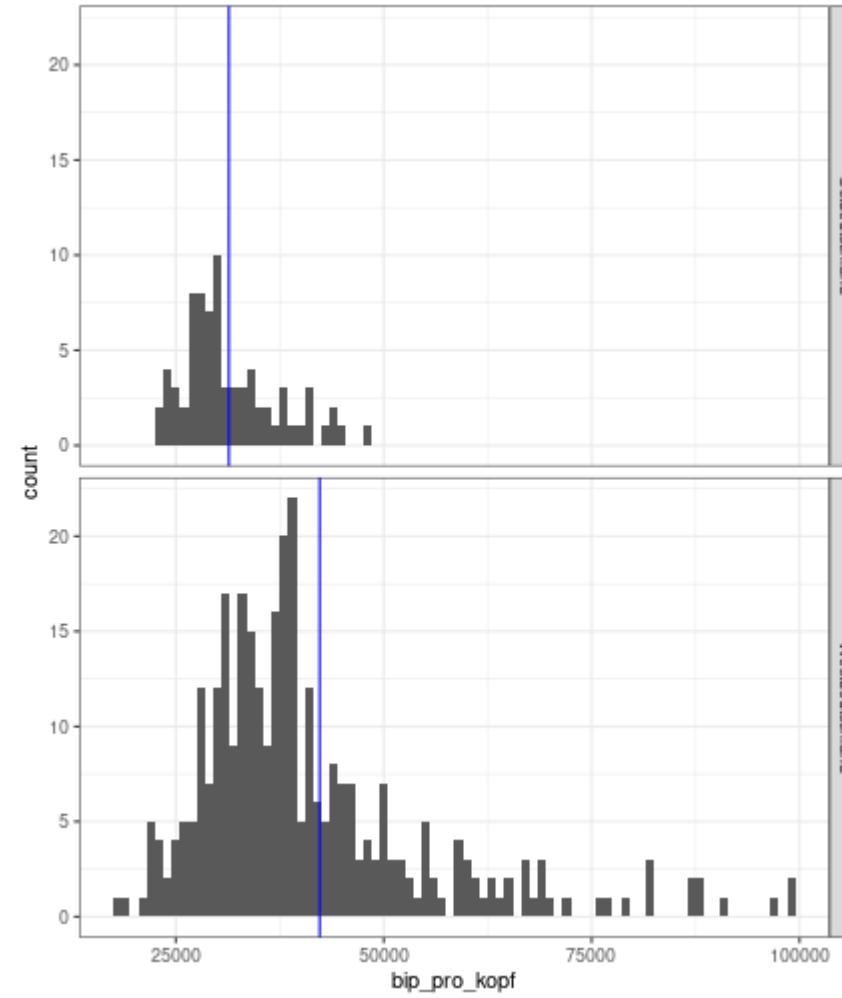
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf > 0))
  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, bip_pro_kopf < 100000),
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
  facet_grid(ost_name ~ .) +
  theme_bw()
```

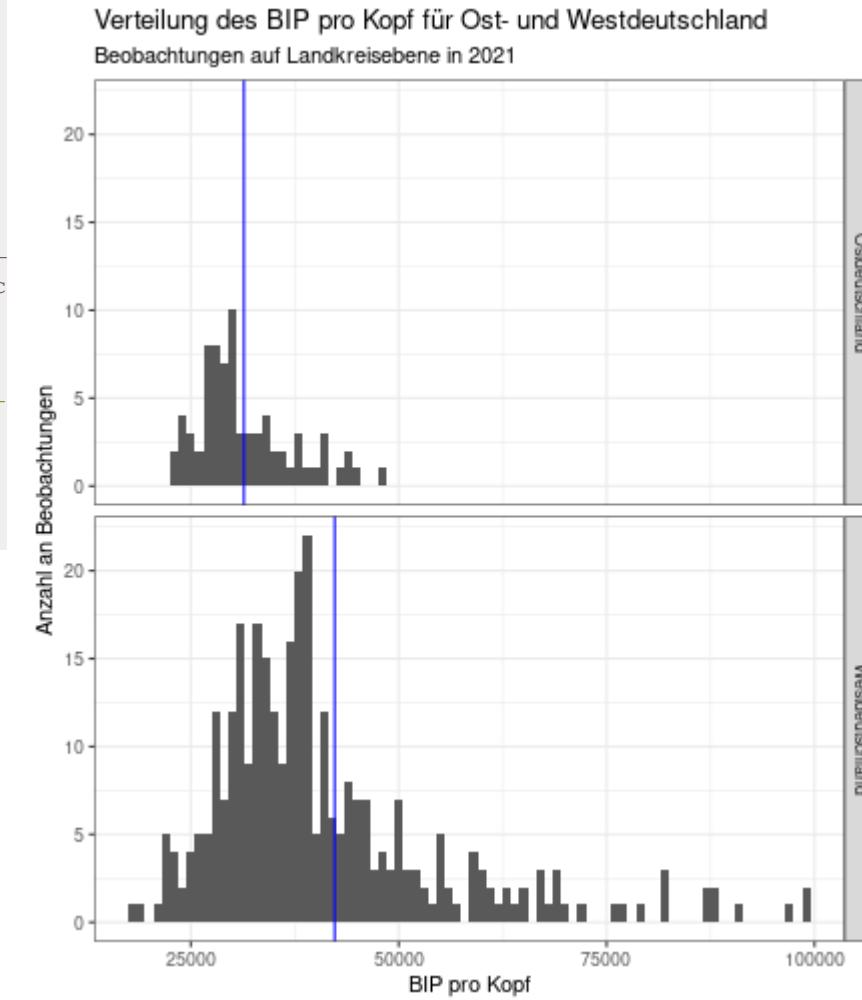


```

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

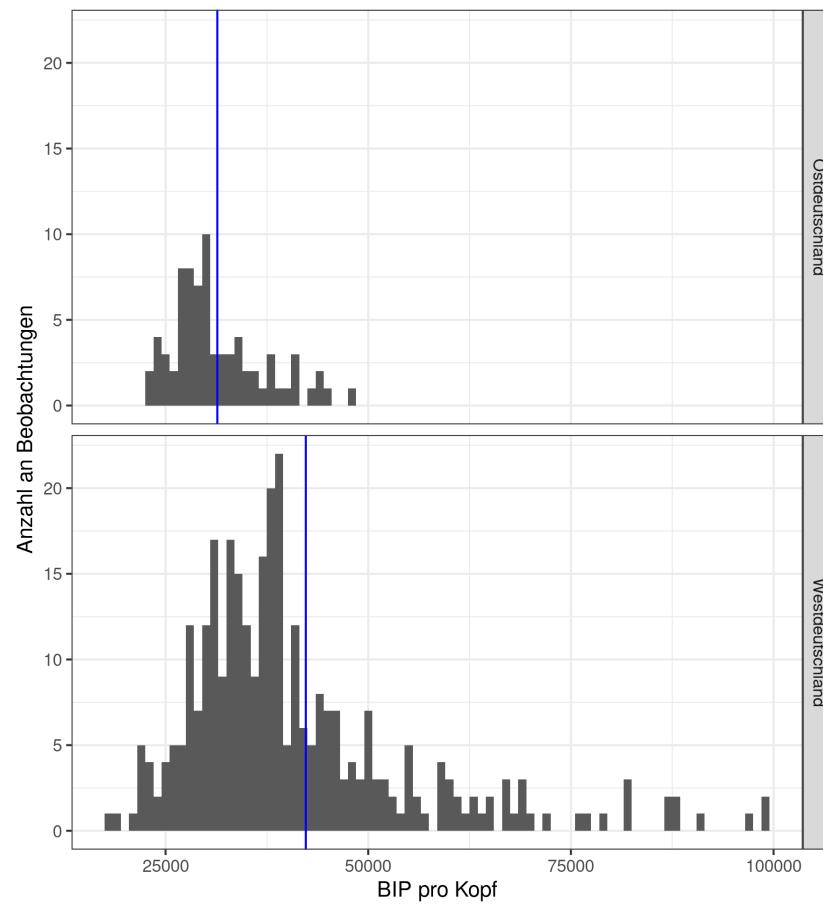
ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf < 100000),
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
  facet_grid(ost_name ~ .) +
  theme_bw() +
  labs(title = "Verteilung des BIP pro Kopf für Ost- und Westdeutschland",
  subtitle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

```



# Verteilung des BIP pro Kopf in 2021

Verteilung des BIP pro Kopf für Ost- und Westdeutschland  
Beobachtungen auf Landkreisebene in 2021



# Verteilung des BIP pro Kopf in 2021

Das Histogramm bestätigen das Bild des Boxplots:

- ✚ Deutliche Unterschiede zwischen ost- und westdeutschend Landkreisen in 2021
- ✚ Deutlich mehr Ausreißer nach oben bei westdeutschen Landkreisen
- ✚ Verteilung ist für ostdeutsche Landkreise enger um den Mittelwert für das BIP pro Kopf von 31384€
- ✚ 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

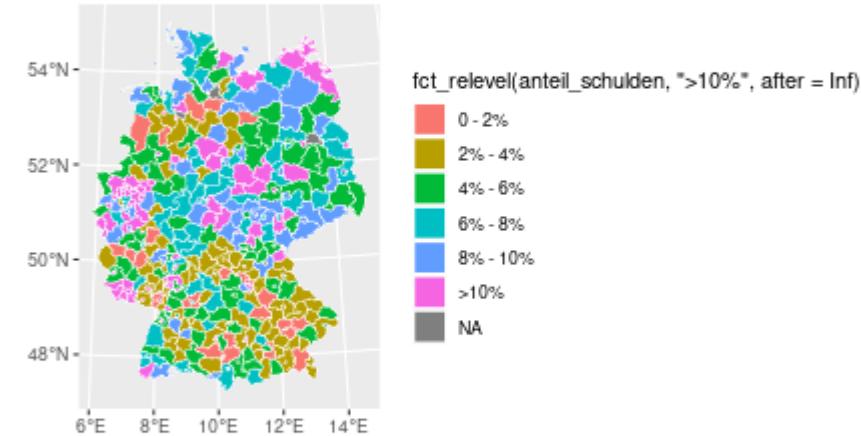
Warum könnte die Verschuldung des öffentlichen Haushalts ein Indikator für eine hohe Arbeitslosenquote sein?

Darstellung der Verschuldung der Landkreise mittels einer Deutschlandkarte.

Beschreiben und interpretieren Sie die folgende Grafik.

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

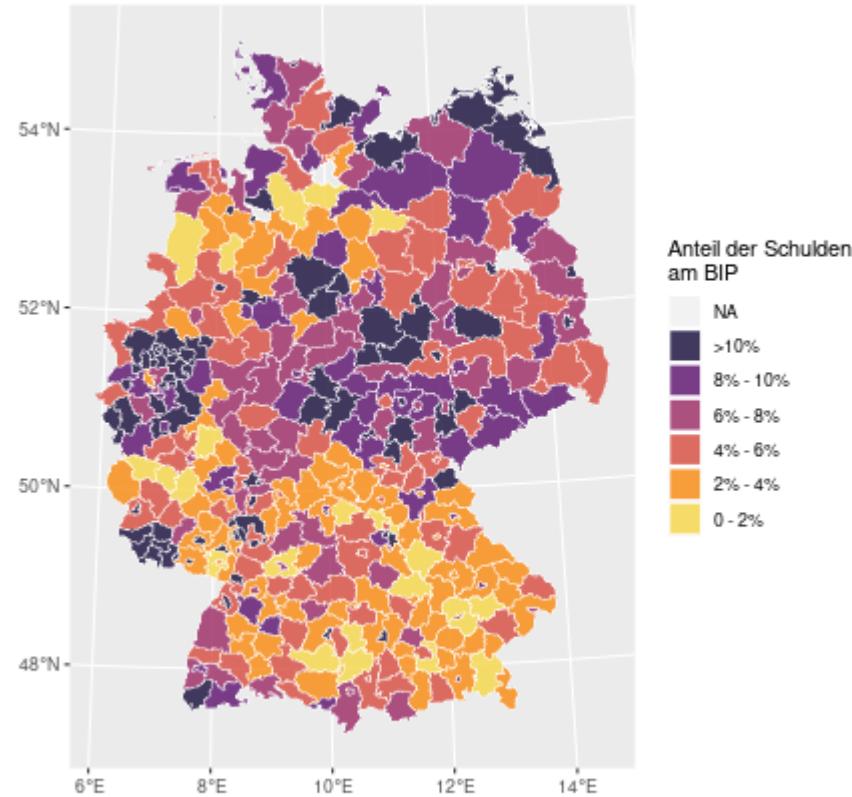
```
ggplot(  
  # Datensatz  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", after = Inf)  
    ),  
    color = "white",  
    size = 0.1  
)
```



```

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))

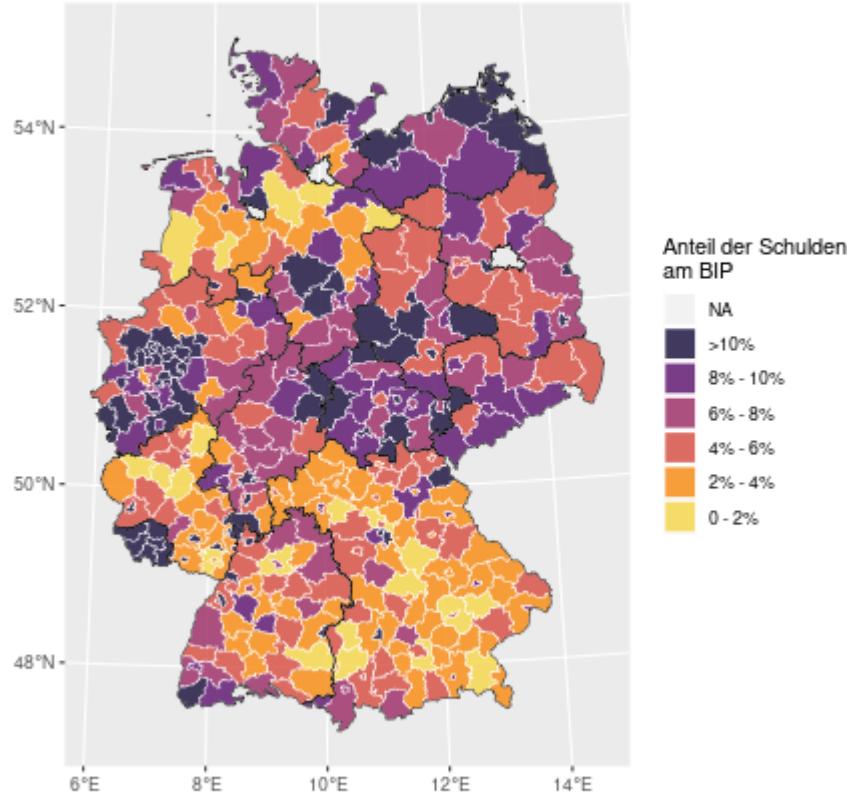
```



```

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 = "black",
      size = 0.5
    )

```

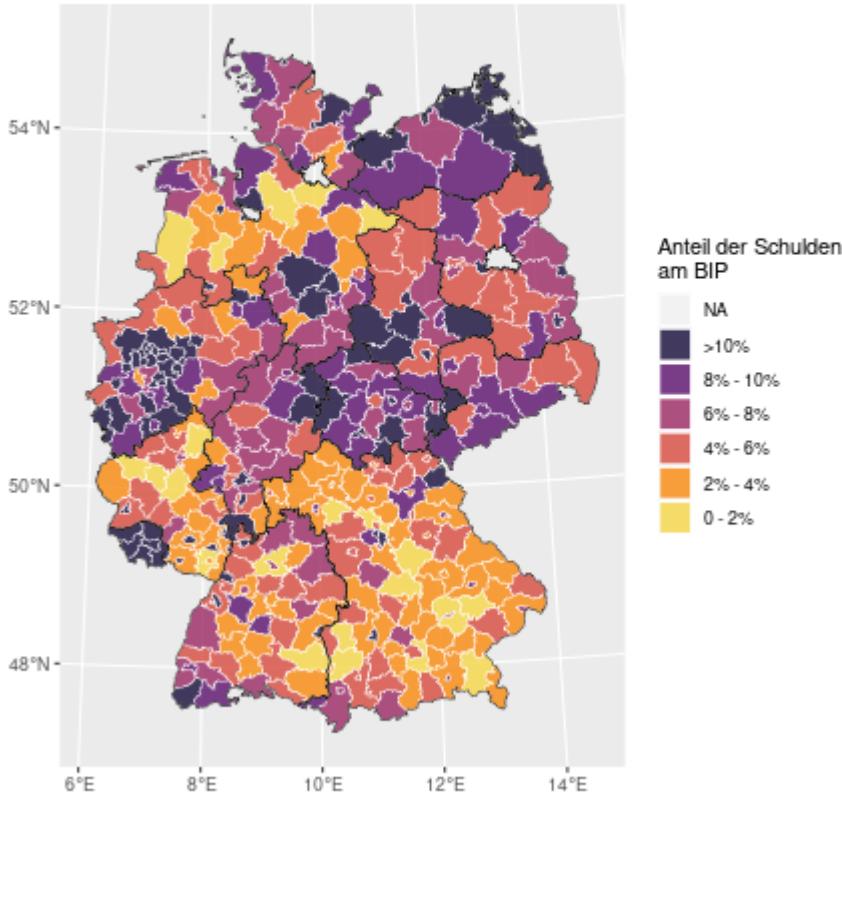


```

ggplot(
  # Datensatz
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", after = 1),
      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 = "black",
      size = 0.5
    ) +
    # Titel
    labs(x = NULL,
         y = NULL,
         title = "Wie verschuldet sind die deutschen Landkreise?",
         subtitle = "Öffentliche Schulden im Vergleich zum BIP in 2021")
  )

```

Wie verschuldet sind die deutschen Landkreise?  
Öffentliche Schulden im Vergleich zum BIP in 2021



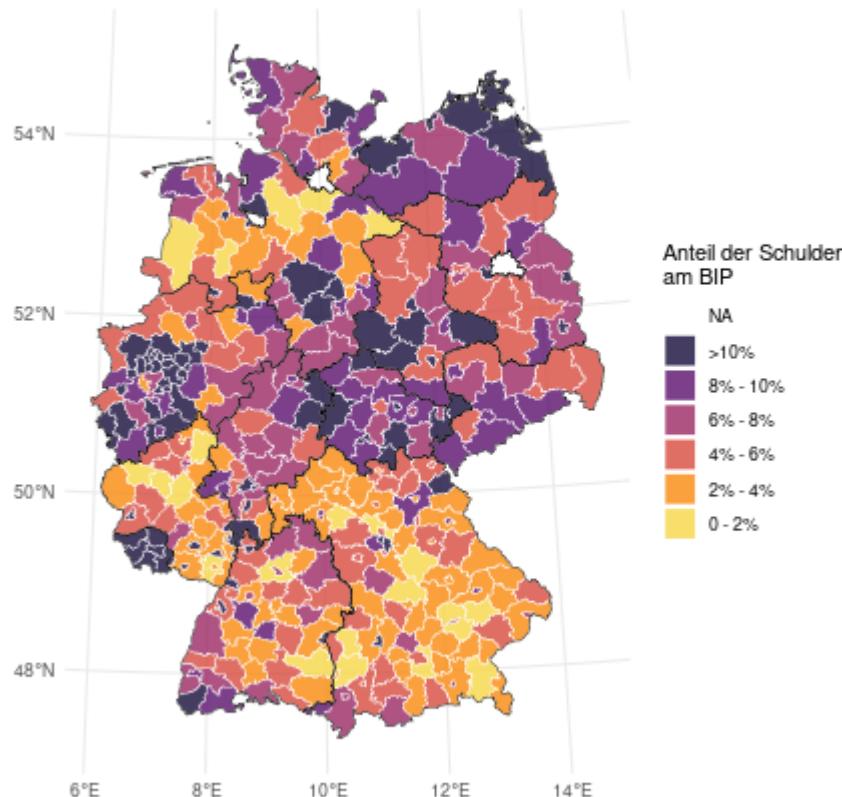
```
ggplot(  
  # Datensatz  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af  
    ),  
    color = "white",  
    size = 0.1  
) +  
  # Viridis Farbschema  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "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 = "black",  
    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%", after = 1),
      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 = TRUE)) +
    # etwas dickere Linien für Bundeslandgrenzen
    geom_sf(
      data = bundesland,
      fill = "transparent",
      color = "black",
      size = 0.5
    ) +
    # Titel
    labs(x = NULL,
         y = NULL,
         title = "Wie verschuldet sind die deutschen Landkreise?",
         subtitle = "Öffentliche Schulden im Vergleich zum BIP in 2021",
         theme_minimal() -> plot_schulden_lk
  )

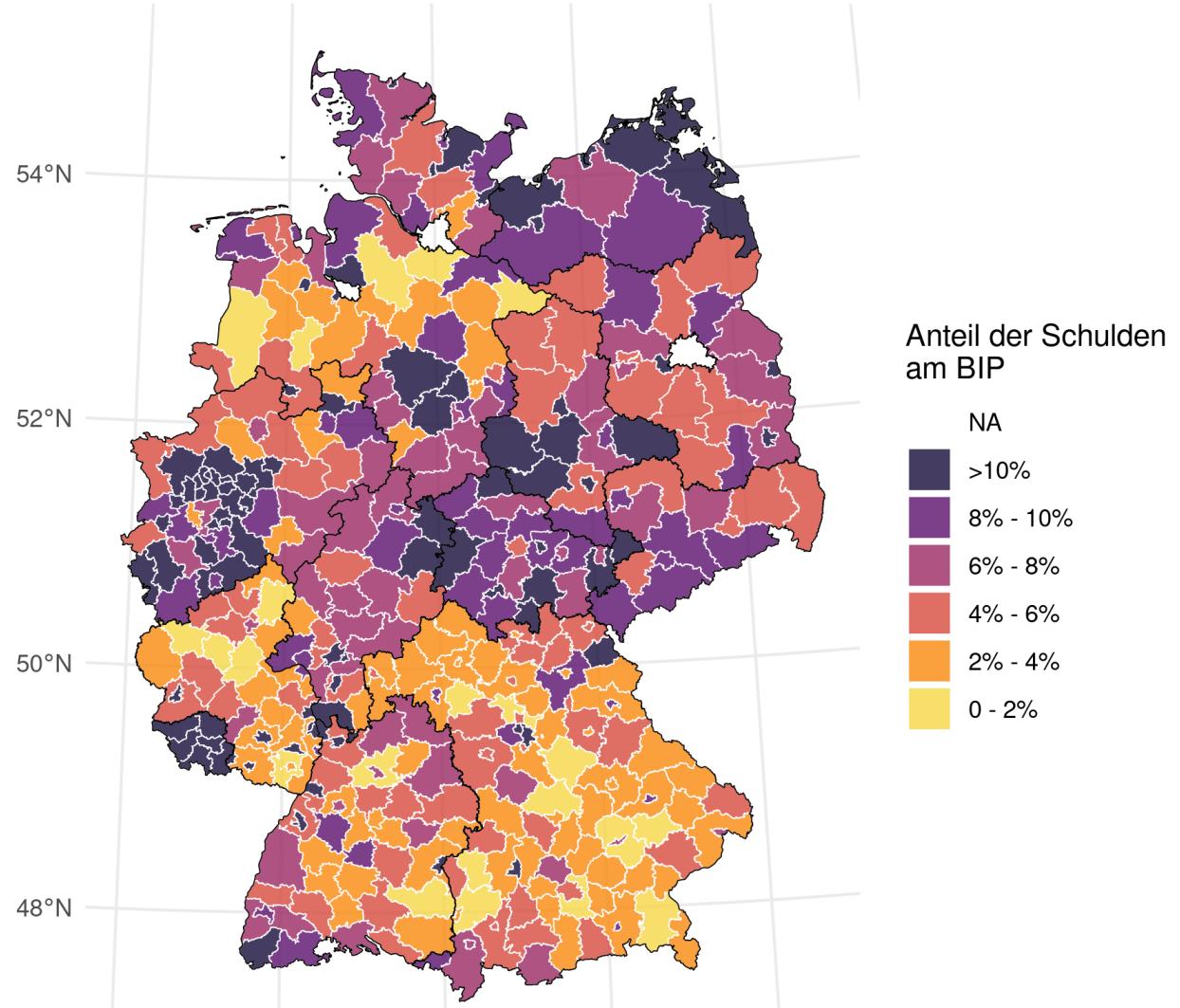
```

Wie verschuldet sind die deutschen Landkreise?  
Öffentliche Schulden im Vergleich zum BIP in 2021



## Wie verschuldet sind die deutschen Landkreise?

Öffentliche Schulden im Vergleich zum BIP in 2021



# Verschuldung

## Beschreibung:

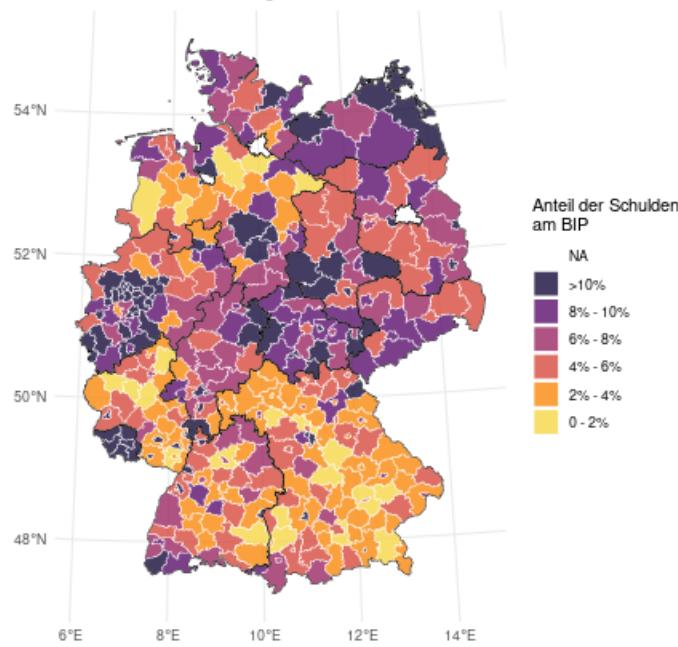
- ✚ Niedrige Verschuldung im Verhältnis zum BIP: Baden-Württemberg, Bayern, Sachsen und Niedersachsen
- ✚ Mittlere Verschuldung: Rheinland-Pfalz, Brandenburg, Hessen, Schleswig-Holstein
- ✚ Hohe Verschuldung: Sachsen-Anhalt, Thüringen, Nordrhein-Westfalen, Saarland, Mecklenburg-Vorpommern

## Interpretation:

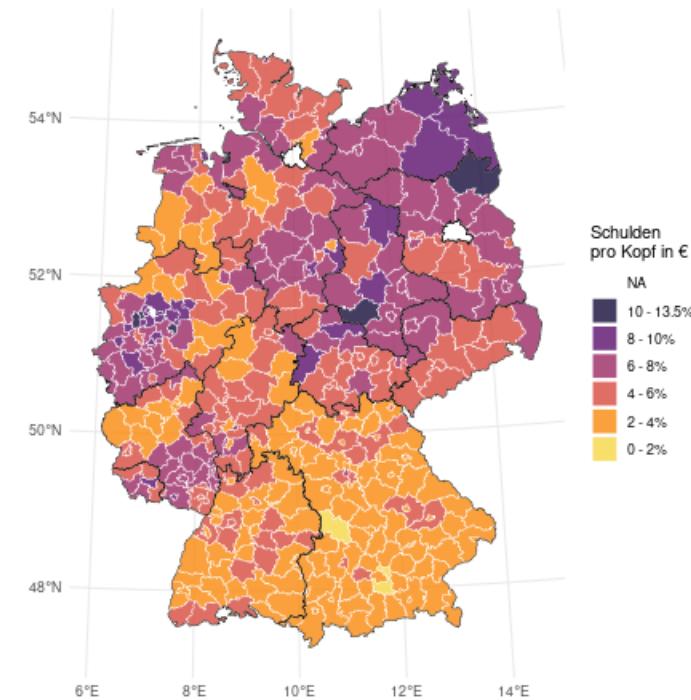
- ✚ Strukturschwache Landkreise sind vermehrt in Ostdeutschland zu finden, allerdings scheint es eher ein Nord/Süd Gefälle als ein Ost/West Gefälle zu geben
- ✚ Die ehemalige Herzammer der deutschen Industrie, das Ruhrgebiet, leidet unter dem Strukturwandel hin zu erneuerbaren Energien
  - ✚ Es fallen 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 2021



Arbeitslosigkeit in Deutschland  
Dargestellt ist die Arbeitslosenquote für alle Landkreise in 2021



# 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  $\mu_x, \mu_y$  als Mittelwerte von  $x_1, \dots, x_n$  und  $y_1, \dots, y_n$ .  $\sigma_x, \sigma_y$  sind die Standardabweichungen von diesem Mittelwert.  $\rho$  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  $\mu_x, \mu_y$  als Mittelwerte von  $x_1, \dots, x_n$  und  $y_1, \dots, y_n$ .  $\sigma_x, \sigma_y$  sind die Standardabweichungen von diesem Mittelwert.  $\rho$  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

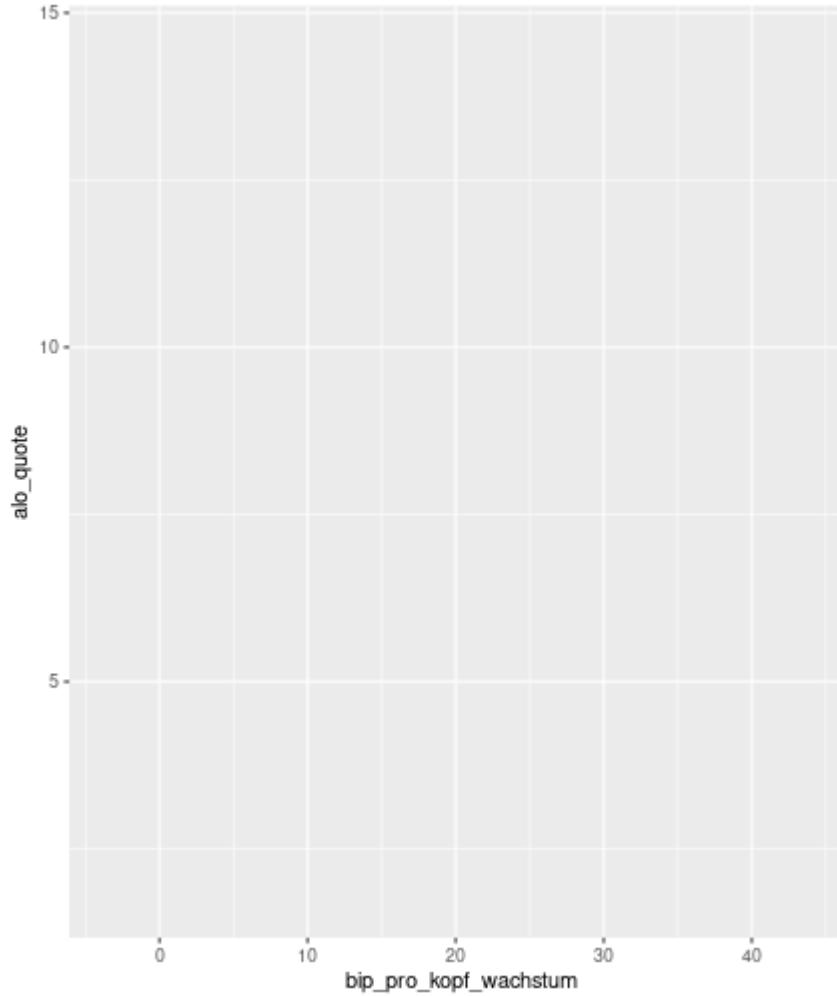
Im Streudiagramm können wir Variable x auf der x-Achse und Variable y auf der y-Achse abtragen

gesamtdaten

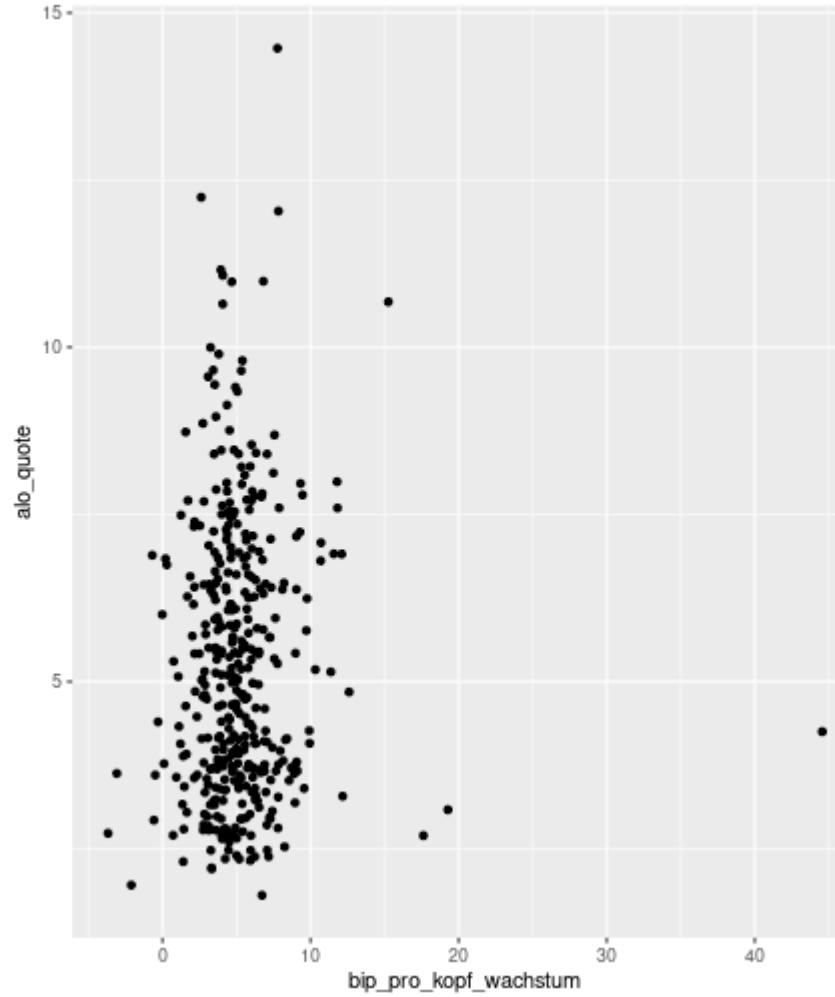
```
## # A tibble: 400 × 17
##   Regionalschluessel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                4369. Flensburg            01        Schleswig-Holst
## 2 01002                11097. Kiel                  01        Schleswig-Holst
## 3 01003                9347. Lübeck                01        Schleswig-Holst
## 4 01004                3771. Neumünster            01        Schleswig-Holst
## 5 01051                4143. Dithmarschen            01        Schleswig-Holst
## 6 01053                5603. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                4699 Nordfriesland            01        Schleswig-Holst
## 8 01055                5371 Ostholstein              01        Schleswig-Holst
## 9 01056                9371. Pinneberg                01        Schleswig-Holst
## 10 01057                2854. Plön                  01        Schleswig-Holst
## # ... i 390 more rows
## # ... i 12 more variables: Schulden_pro_kopf_1k <dbl>, Einwohner <dbl>,
## #       Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #       alo_quote <dbl>, ost <fct>, ost_name <chr>, bip_pro_kopf_wachstum <dbl>,
## #       Jahr <dbl>, anteil_schulden <dbl>
```

```
gesamtdataen %>%
```

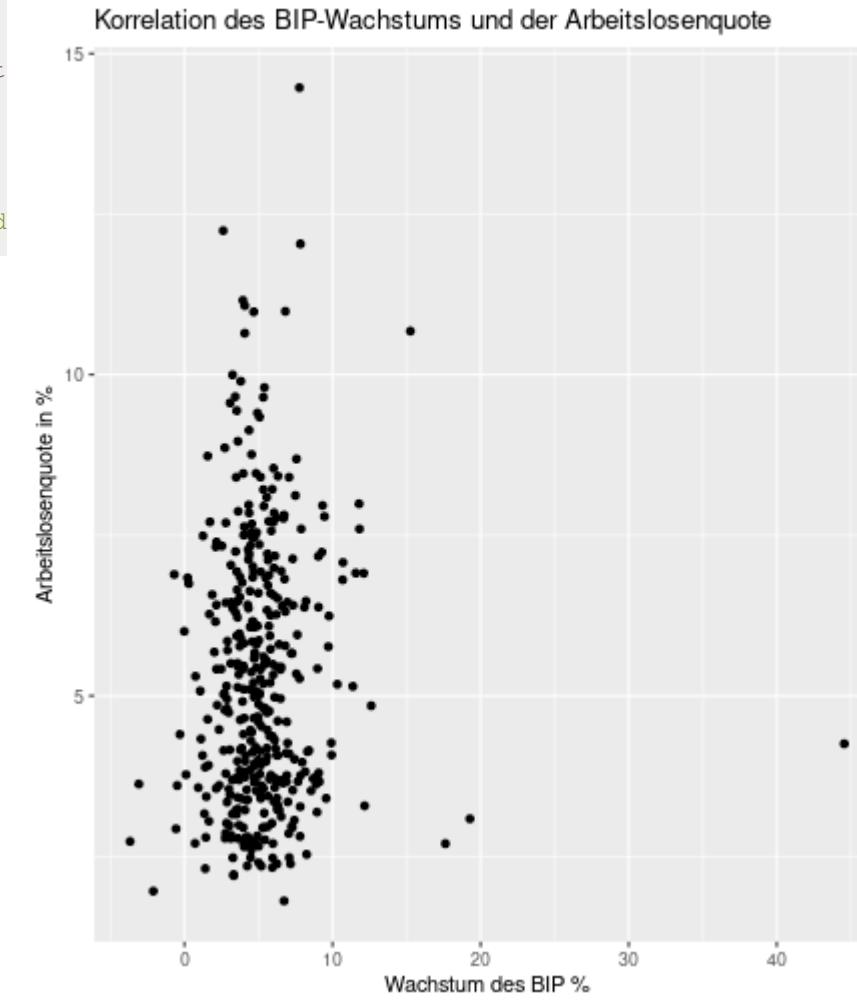
```
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot
```



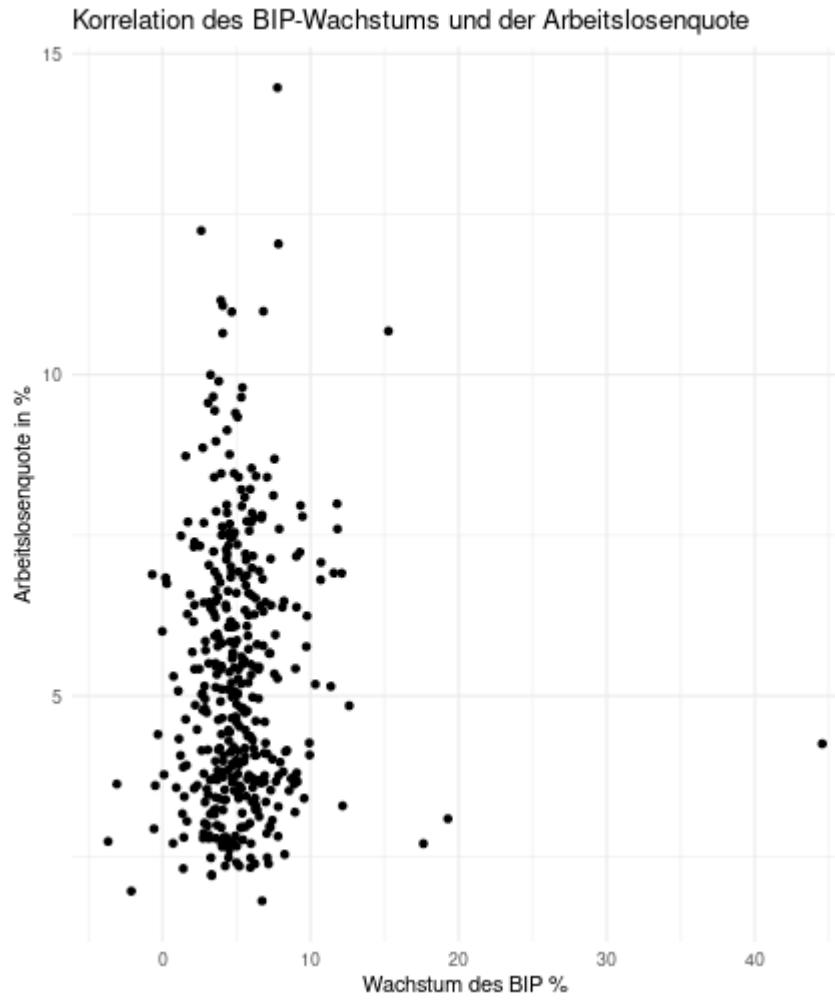
```
gesamtdaten %>%  
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot  
  geom_point()
```



```
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
```



```
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 zwischen Arbeitslosenquote und BIP-Wachstum

- ✚ Es fallen die Ausreißer ins Auge (+40% und -4%)
  - ✚ Vorheriges Jahr hohes/niedriges BIP, dadurch jetzt niedriges/hohes BIP-Wachstum
  - ✚ Corona war beides mal der Auslöser (Mainz und Wolfsburg die betroffenen Landkreise)
- ✚ 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] 0.02804881
```

# Korrelation zwischen Arbeitslosenquote und BIP-Wachstum

- ✚ Es fallen die Ausreißer ins Auge (+40% und -4%)
  - ✚ Vorheriges Jahr hohes/niedriges BIP, dadurch jetzt niedriges/hohes BIP-Wachstum
  - ✚ Corona war beides mal der Auslöser (Mainz und Wolfsburg die betroffenen Landkreise)
- ✚ 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")
```

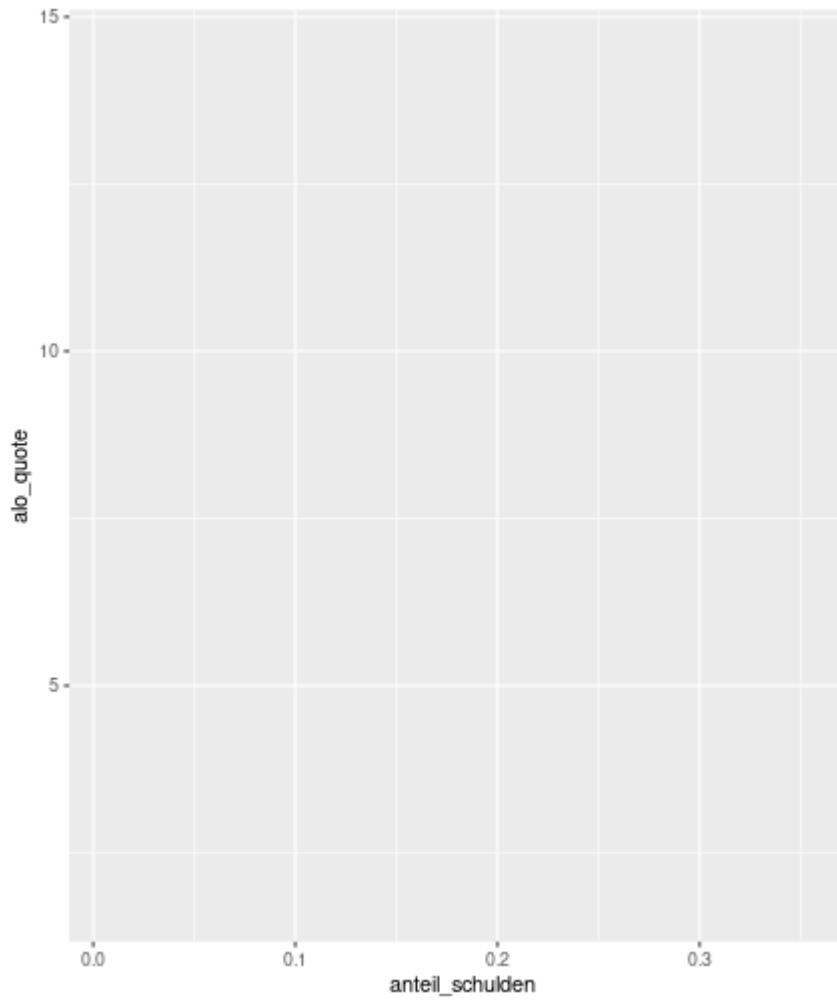
```
## [1] 0.02804881
```

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

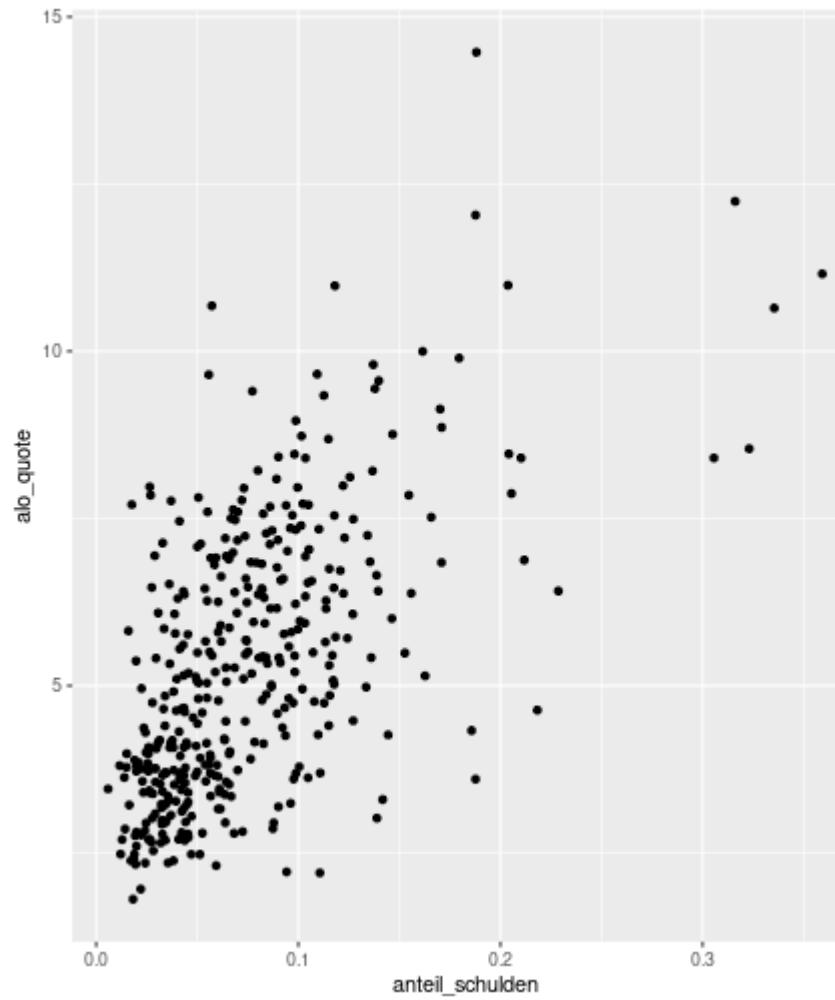
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g
```

```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten  
## # A tibble: 400 × 17  
## # Regionschlüssel total_alo landkreis_name      bundesland bundesland_name  
## # <chr>          <dbl> <chr>          <chr>      <chr>  
## # 1 01001          4369. Flensburg          01       Schleswig-Holst  
## # 2 01002          11097. Kiel            01       Schleswig-Holst  
## # 3 01003          9347. Lübeck          01       Schleswig-Holst  
## # 4 01004          3771. Neumünster        01       Schleswig-Holst  
## # 5 01051          4143. Dithmarschen        01       Schleswig-Holst  
## # 6 01053          5603. Herzogtum Lauenburg 01       Schleswig-Holst  
## # 7 01054          4699 Nordfriesland        01       Schleswig-Holst  
## # 8 01055          5371 Ostholstein        01       Schleswig-Holst  
## # 9 01056          9371. Pinneberg        01       Schleswig-Holst  
## # 10 01057         2854. Plön            01       Schleswig-Holst  
## # i 390 more rows  
## # i 12 more variables: Schulden_pro_kopf_1k <dbl>, Einwohner <dbl>,  
## #   Schulden_gesamt <dbl>, bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,  
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>, bip_pro_kopf_wachstum <dbl>,  
## #   Jahr <dbl>, anteil_schulden <dbl>
```

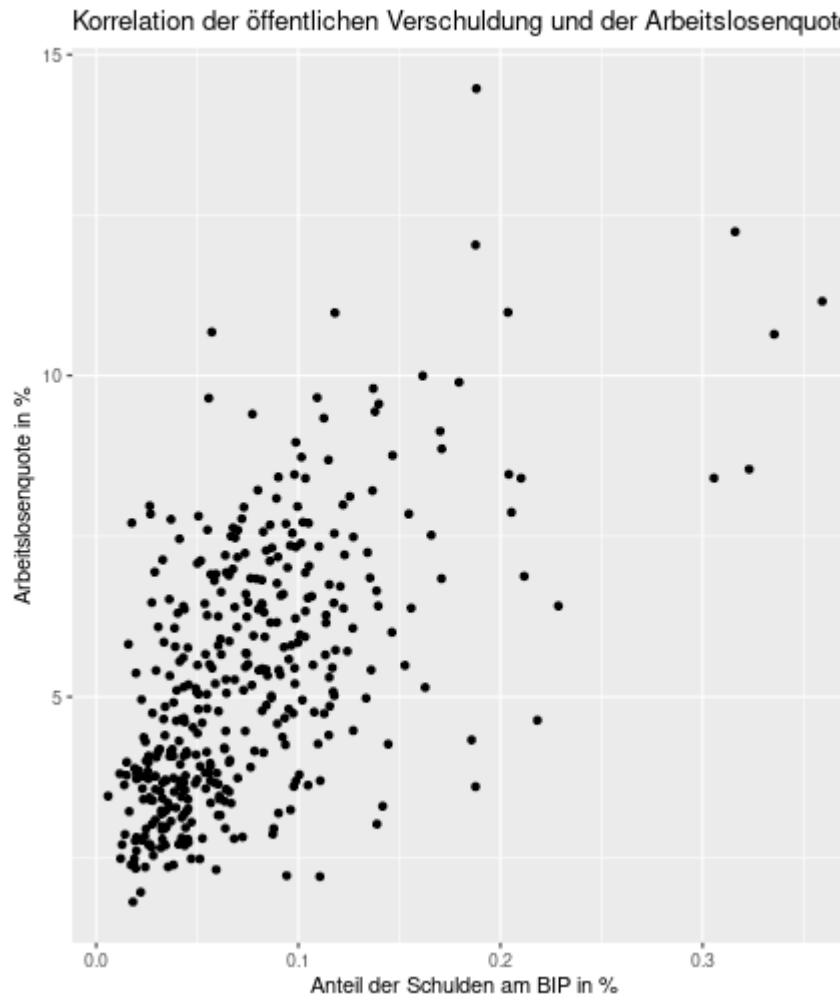
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote))
```



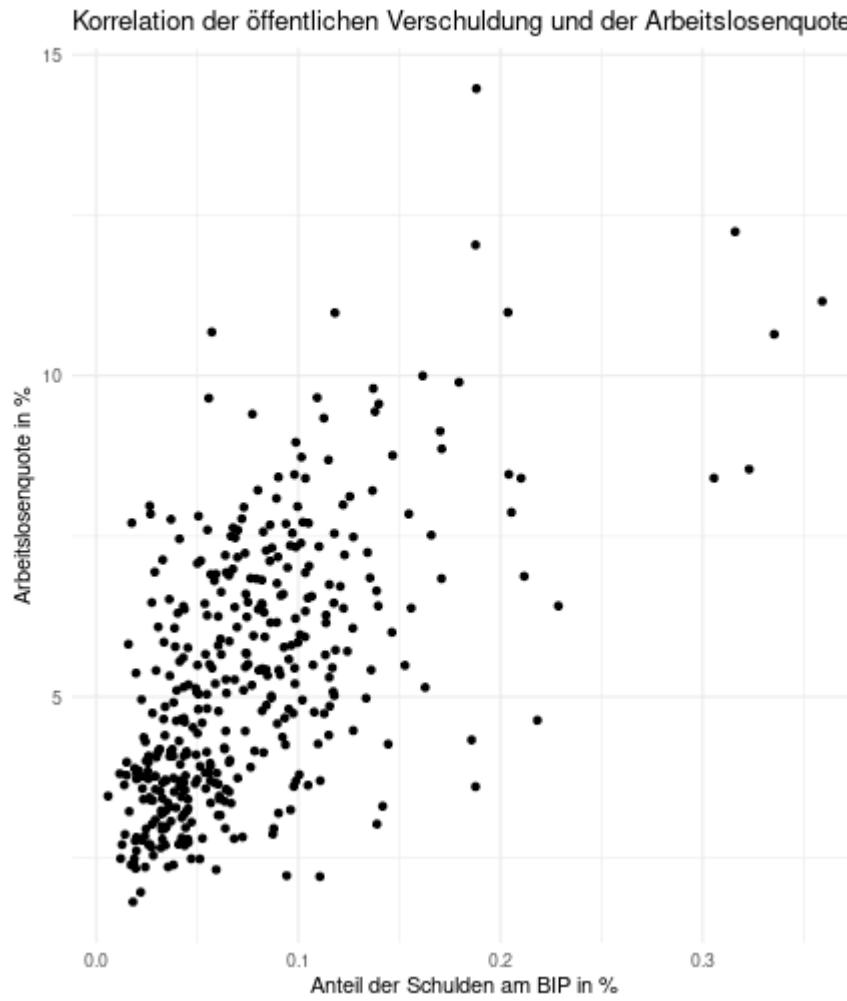
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point()
```



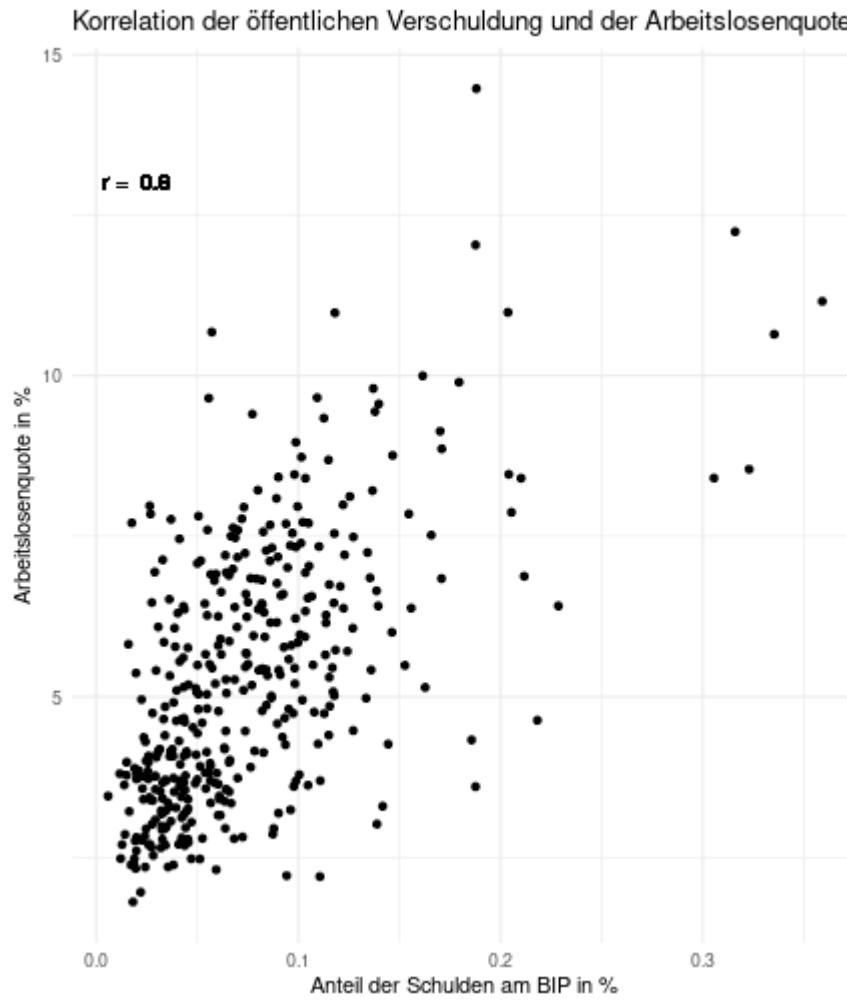
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Anteil der Schulden am BIP in %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation der öffentlichen Versch"
```



```
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  $\rho = 0.60$  auch einen starken Zusammenhang

$\rho$  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  $\rho = 0.60$  auch einen starken Zusammenhang

$\rho$  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 2021 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.