

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
- + Arbeiten mit AI → Lernen wie AI ihre Arbeit sinnvoll ergänzen kann und wo nicht

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 oder `lapply`
 - + 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

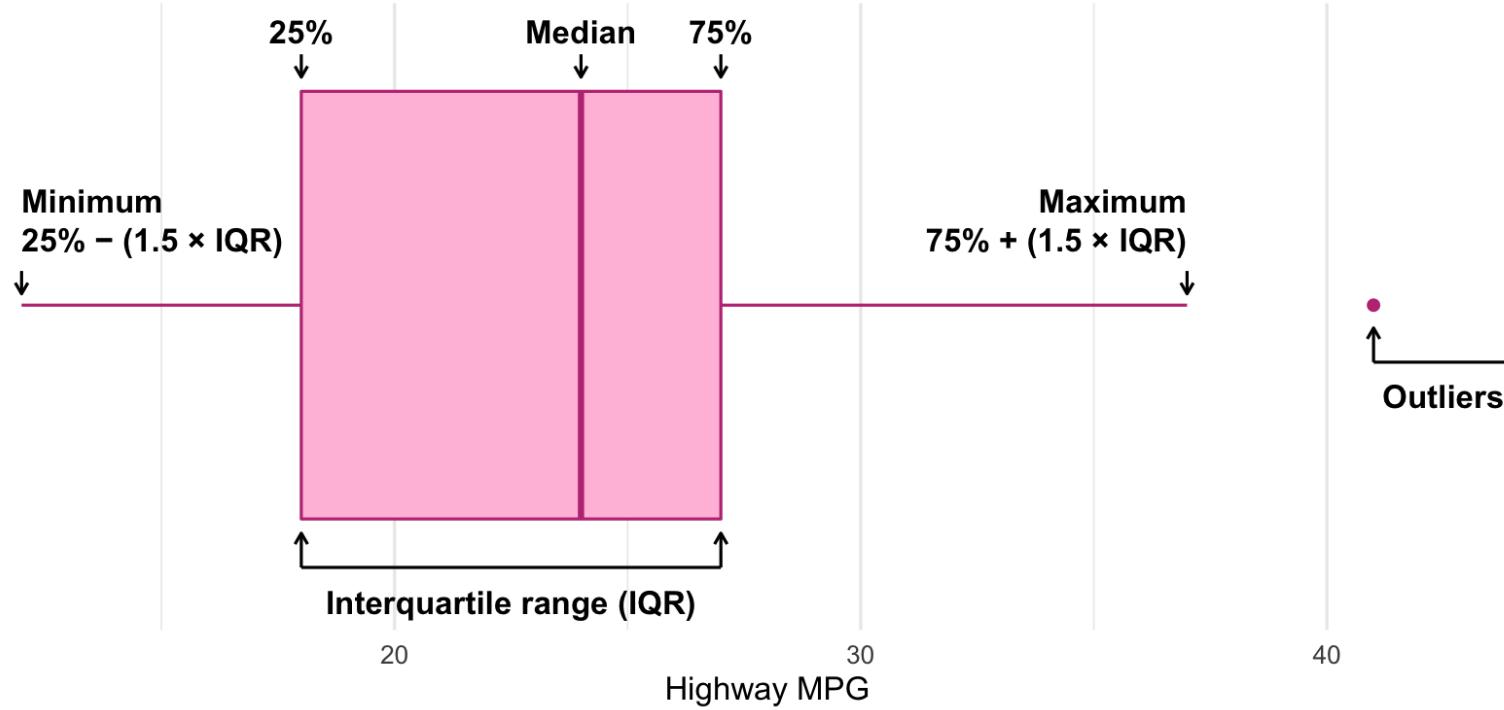
Univariate deskriptive Statistik

- + 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 erstellen.

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    42115549        77798888
```

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    42115549        77798888
```

- +
- 42,1 Mio. Erwerbstätige und 77,8 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

Table: 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	4.89	2.01	1.50	3.30	4.66	6.12	13.7
Schulden_pro_kopf_lk	4	0.99	3002.65	2300.05	218.11	1492.76	2338.12	3635.14	17032.2
bip_pro_kopf	2	1.00	42801.26	17280.40	17953.34	32377.34	38778.96	46569.47	149442.9
Einwohner	4	0.99	196461.84	153389.80	34426.00	104450.75	156913.50	241289.50	1508933.0

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	4.89	2.01	1.50	3.30	4.66	6.12	13.71
Schulden_pro_kopf_lk	4	0.99	3002.65	2300.05	218.11	1492.76	2338.12	3635.14	17032.20
bip_pro_kopf	2	1.00	42801.26	17280.40	17953.34	32377.34	38778.96	46569.47	149442.98

Beschreibung der Tabelle

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
alo_quote	2	1.00	4.89	2.01	1.50	3.30	4.66	6.12	13.71
Schulden_pro_kopf_lk	4	0.99	3002.65	2300.05	218.11	1492.76	2338.12	3635.14	17032.20
bip_pro_kopf	2	1.00	42801.26	17280.40	17953.34	32377.34	38778.96	46569.47	149442.98

Bitte diskutieren Sie mit ihrem/ihrer Sitznachbar*in und 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 die folgenden Punkte ein:

- + Mittelwert
- + Standardabweichung
- + Median

05 : 00

Arbeitslosenquote

Mittelwert: 4,89 Prozent

- + Hoch
- + Jedoch SGB II und SGB III
- + Konsistenzcheck auf [Statista](#) zeigt eine Arbeitslosenquote von 5,3% für 2022
- + **Jedoch:** Wir haben nicht Berlin und Hamburg in den Daten

Standardabweichung: 2,01

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

Median: 4,66 Prozent

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

Verschuldung pro Kopf

Mittelwert: 3003€

- + Relativ hoch

Standardabweichung: 2300€

- + Sehr hohe Streuung
- + Deutliche regionale Unterschiede

Median: 2338€

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

BIP pro Kopf

Mittelwert: 42801€

- + Insgesamt recht hoch
- + Starker Wirtschaftsstandort Deutschland

Standardabweichung: 17280€

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

Median: 38779€

- + 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 2022
- ✚ Alle Landkreise
- ✚ Für einige Landkreise haben wir keine Informationen (sogenannte "Missing values" -> n_missing)

Wir möchten nun die regionale Verteilung der Arbeitslosenquote in Deutschland im Jahr 2022 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), .groups = 'dr
```

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            3970. Flensburg        01       Schleswig-Holst
## 2 01002            10315. Kiel           01       Schleswig-Holst
## 3 01003            8776. Lübeck         01       Schleswig-Holst
## 4 01004            3359. Neumünster     01       Schleswig-Holst
## 5 01051            3858. Dithmarschen    01       Schleswig-Holst
## 6 01053            5351. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4155. Nordfriesland    01       Schleswig-Holst
## 8 01055            4824. Ostholstein     01       Schleswig-Holst
## 9 01056            8547. Pinneberg       01       Schleswig-Holst
## 10 01057           2572. 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 )
```

```
## # A tibble: 400 × 12
## # Groups:   bundesland_name [16]
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl>   <chr>          <chr>        <chr>
## 1 01001            3970. Flensburg             01 Schleswig-Holst
## 2 01002            10315. Kiel                  01 Schleswig-Holst
## 3 01003            8776. Lübeck                01 Schleswig-Holst
## 4 01004            3359. Neumünster            01 Schleswig-Holst
## 5 01051            3858. Dithmarschen           01 Schleswig-Holst
## 6 01053            5351. Herzogtum Lauenburg    01 Schleswig-Holst
## 7 01054            4155. Nordfriesland          01 Schleswig-Holst
## 8 01055            4824. Ostholstein            01 Schleswig-Holst
## 9 01056            8547. Pinneberg              01 Schleswig-Holst
## 10 01057           2572. 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), .groups
```

Die Arbeitslosenquote auf Bundeslandebene

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

```
## # A tibble: 14 × 4
##   bundesland_name     mean_alo   sd_alo   median
##   <chr>          <dbl>    <dbl>    <
## 1 Bayern           2.98    0.703
## 2 Baden-Württemberg 3.41    0.741
## 3 Hessen            4.72    1.17
## 4 Rheinland-Pfalz  5.05    1.44
## 5 Schleswig-Holstein 5.38    0.865
## 6 Sachsen           5.53    0.790
## 7 Niedersachsen    5.57    1.67
## 8 Saarland          5.57    1.67
## 9 Thüringen         5.64    1.35
## 10 Brandenburg      6.36    1.42
## 11 Nordrhein-Westfalen 6.63    2.37
## 12 Mecklenburg-Vorpommern 7.21    1.17
## 13 Sachsen-Anhalt    7.45    1.38
## 14 Bremen            9.08    2.64
```

Die Arbeitslosenquote auf Bundeslandebene

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

```
## # A tibble: 14 × 4
##   bунdesland_name    mean_alo   sd_alo median
##   <chr>          <dbl>     <dbl>   <dbl>
## 1 Bayern           2.98      0.703   2.98
## 2 Baden-Württemberg 3.41      0.741   3.41
## 3 Hessen            4.72      1.17    4.72
## 4 Rheinland-Pfalz  5.05      1.44    5.05
## 5 Schleswig-Holstein 5.38      0.865   5.38
## 6 Sachsen           5.53      0.790   5.53
## 7 Niedersachsen    5.57      1.67    5.57
## 8 Saarland          5.57      1.67    5.57
## 9 Thüringen         5.64      1.35    5.64
## 10 Brandenburg      6.36      1.42    6.36
## 11 Nordrhein-Westfalen 6.63      2.37    6.63
## 12 Mecklenburg-Vorpommern 7.21      1.17    7.21
## 13 Sachsen-Anhalt   7.45      1.38    7.45
## 14 Bremen            9.08      2.64    9.08
```

Bundesland	Arbeitslosenquote		
	Mittelwert	Std.	Median
Bayern	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

Bitte beachten:

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

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Die Arbeitslosenquote auf Bundeslandebene

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

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

```
bulu_data %>%
  arrange( mean_alo ) %>%
  filter( !is.na(mean_alo) ) %>%
  kbl(col.names = c("Bundesland",
                    "Mittelwert",
                    "Std.",
                    "Median"), digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive")) %>%
  kable_paper(full_width = F) %>%
  row_spec(c(6,9, 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 3.41     0.741     3.31
## 2 Bayern            2.98     0.703     2.88
## 3 Berlin             NA       NA        NA
## 4 Brandenburg       6.36     1.42      6.61
## 5 Bremen             9.08     2.64      9.08
## 6 Hamburg            NA       NA        NA
## 7 Hessen              4.72     1.17      4.83
## 8 Mecklenburg-Vorpommern 7.21     1.17      7.41
## 9 Niedersachsen      5.57     1.67      5.75
## 10 Nordrhein-Westfalen 6.63     2.37      6.26
## 11 Rheinland-Pfalz   5.05     1.44      4.91
## 12 Saarland           5.57     1.67      5.25
## 13 Sachsen            5.53     0.790     5.39
## 14 Sachsen-Anhalt    7.45     1.38      7.23
## 15 Schleswig-Holstein 5.38     0.865     5.46
## 16 Thüringen          5.64     1.35      5.14
```

```
bula_data %>%
```

```
  arrange( mean_alo )
```

	bundesland_name	mean_alo	sd_alo	median_alo
	<chr>	<dbl>	<dbl>	<dbl>
## 1	Bayern	2.98	0.703	2.88
## 2	Baden-Württemberg	3.41	0.741	3.31
## 3	Hessen	4.72	1.17	4.83
## 4	Rheinland-Pfalz	5.05	1.44	4.91
## 5	Schleswig-Holstein	5.38	0.865	5.46
## 6	Sachsen	5.53	0.790	5.39
## 7	Niedersachsen	5.57	1.67	5.75
## 8	Saarland	5.57	1.67	5.25
## 9	Thüringen	5.64	1.35	5.14
## 10	Brandenburg	6.36	1.42	6.61
## 11	Nordrhein-Westfalen	6.63	2.37	6.26
## 12	Mecklenburg-Vorpommern	7.21	1.17	7.41
## 13	Sachsen-Anhalt	7.45	1.38	7.23
## 14	Bremen	9.08	2.64	9.08
## 15	Berlin	NA	NA	NA
## 16	Hamburg	NA	NA	NA

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

```
## # A tibble: 14 × 4
##   bundesland_name      mean_alo    sd_alo median_alo
##   <chr>          <dbl>     <dbl>      <dbl>
## 1 Bayern           2.98     0.703     2.88
## 2 Baden-Württemberg 3.41     0.741     3.31
## 3 Hessen            4.72     1.17      4.83
## 4 Rheinland-Pfalz  5.05     1.44      4.91
## 5 Schleswig-Holstein 5.38     0.865     5.46
## 6 Sachsen           5.53     0.790     5.39
## 7 Niedersachsen    5.57     1.67      5.75
## 8 Saarland          5.57     1.67      5.25
## 9 Thüringen         5.64     1.35      5.14
## 10 Brandenburg     6.36     1.42      6.61
## 11 Nordrhein-Westfalen 6.63     2.37      6.26
## 12 Mecklenburg-Vorpommern 7.21     1.17      7.41
## 13 Sachsen-Anhalt   7.45     1.38      7.23
## 14 Bremen            9.08     2.64      9.08
```

```

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	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

```

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	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

```

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)

```

Bundesland	Mittelwert	Std.	Median
Bayern	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

```

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(6, 9, 10, 12, 13), bold = T, color = "whi

```

Bundesland	Mittelwert	Std.	Median
Bayern	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

```

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(6, 9, 10, 12, 13), bold = T, color = "whi
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Arbeitslosenquote				
Bundesland	Mittelwert	Std.	Median	
Bayern	2.98	0.70	2.88	
Baden-Württemberg	3.41	0.74	3.31	
Hessen	4.72	1.17	4.83	
Rheinland-Pfalz	5.05	1.44	4.91	
Schleswig-Holstein	5.38	0.86	5.46	
Sachsen	5.53	0.79	5.39	
Niedersachsen	5.57	1.67	5.75	
Saarland	5.57	1.67	5.25	
Thüringen	5.64	1.35	5.14	
Brandenburg	6.36	1.42	6.61	
Nordrhein-Westfalen	6.63	2.37	6.26	
Mecklenburg-Vorpommern	7.21	1.17	7.41	
Sachsen-Anhalt	7.45	1.38	7.23	
Bremen	9.08	2.64	9.08	

```

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(6, 9, 10, 12, 13), bold = T, color = "whi
  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	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

Bitte beachten:

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

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Was ist hier eine Beobachtung?

Bundesland	Arbeitslosenquote		
	Mittelwert	Std.	Median
Bayern	2.98	0.70	2.88
Baden-Württemberg	3.41	0.74	3.31
Hessen	4.72	1.17	4.83
Rheinland-Pfalz	5.05	1.44	4.91
Schleswig-Holstein	5.38	0.86	5.46
Sachsen	5.53	0.79	5.39
Niedersachsen	5.57	1.67	5.75
Saarland	5.57	1.67	5.25
Thüringen	5.64	1.35	5.14
Brandenburg	6.36	1.42	6.61
Nordrhein-Westfalen	6.63	2.37	6.26
Mecklenburg-Vorpommern	7.21	1.17	7.41
Sachsen-Anhalt	7.45	1.38	7.23
Bremen	9.08	2.64	9.08

Bitte beachten:

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

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine eher niedrige Arbeitslosenquote (<3.5%)

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine eher niedrige Arbeitslosenquote (<3.5%)
- + Landkreise in den ostdeutschen Bundesländern haben tendenziell höhere Arbeitslosenquoten (>5.5%)

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine eher niedrige Arbeitslosenquote (<3.5%)
- + Landkreise in den ostdeutschen Bundesländern haben tendenziell höhere Arbeitslosenquoten (>5.5%)
- + Bundesländer wie Bayern oder Baden-Württemberg haben relativ niedrige Arbeitslosenquoten und relativ geringe regionale Unterschiede (niedrige Standardabweichungen).

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine eher niedrige Arbeitslosenquote (<3.5%)
- + Landkreise in den ostdeutschen Bundesländern haben tendenziell höhere Arbeitslosenquoten (>5.5%)
- + Bundesländer wie Bayern oder Baden-Württemberg haben relativ niedrige Arbeitslosenquoten und relativ geringe regionale Unterschiede (niedrige Standardabweichungen).
- + Bundesländer wie Nordrhein-Westfalen und Niedersachsen haben zwar eine relativ niedrige Arbeitslosenquote, aber sehr unterschiedliche regionale Verteilungen (hohe Standardabweichungen), was auf starke regionale Unterschiede hinweist.

Die Arbeitslosenquote auf Bundeslandebene

Was lernen wir aus der deskriptiven Tabelle?

- + Landkreise im Süden Deutschlands haben durchschnittlich eine eher niedrige Arbeitslosenquote (<3.5%)
- + Landkreise in den ostdeutschen Bundesländern haben tendenziell höhere Arbeitslosenquoten (>5.5%)
- + Bundesländer wie Bayern oder Baden-Württemberg haben relativ niedrige Arbeitslosenquoten und relativ geringe regionale Unterschiede (niedrige Standardabweichungen).
- + Bundesländer wie Nordrhein-Westfalen und Niedersachsen haben zwar eine relativ niedrige Arbeitslosenquote, aber sehr unterschiedliche regionale Verteilungen (hohe Standardabweichungen), was auf starke regionale Unterschiede hinweist.
- + Median liegt recht nahe am Mittelwert für die Bundesländer

Sehr große Unterschiede in den durchschnittlichen Arbeitslosenquoten zwischen Landkreisen in Ost- und Westdeutschland!

Die Arbeitslosenquote zwischen Ost- und Westdeutschland

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

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

gesamtdaten

```
## # A tibble: 400 × 12
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  3970. Flensburg          01        Schleswig-Holst
## 2 01002                  10315. Kiel             01        Schleswig-Holst
## 3 01003                  8776. Lübeck           01        Schleswig-Holst
## 4 01004                  3359. Neumünster       01        Schleswig-Holst
## 5 01051                  3858. Dithmarschen      01        Schleswig-Holst
## 6 01053                  5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4155. Nordfriesland     01        Schleswig-Holst
## 8 01055                  4824. Ostholstein      01        Schleswig-Holst
## 9 01056                  8547. Pinneberg         01        Schleswig-Holst
## 10 01057                 2572. 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                 3970. Flensburg            01        Schleswig-Holst
## 2 01002                 10315. Kiel               01        Schleswig-Holst
## 3 01003                 8776. Lübeck              01        Schleswig-Holst
## 4 01004                 3359. Neumünster         01        Schleswig-Holst
## 5 01051                 3858. Dithmarschen        01        Schleswig-Holst
## 6 01053                 5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                 4155. Nordfriesland       01        Schleswig-Holst
## 8 01055                 4824. Ostholstein        01        Schleswig-Holst
## 9 01056                 8547. Pinneberg           01        Schleswig-Holst
## 10 01057                2572. 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 Berechnur
general_title = "Bitte beachten: ")
```

gesamtdaten

```
## # A tibble: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>          <chr>      <chr>
## 1 01001            3970. Flensburg        01       Schleswig-Holst
## 2 01002            10315. Kiel           01       Schleswig-Holst
## 3 01003            8776. Lübeck         01       Schleswig-Holst
## 4 01004            3359. Neumünster     01       Schleswig-Holst
## 5 01051            3858. Dithmarschen    01       Schleswig-Holst
## 6 01053            5351. Herzogtum Lauenburg 01       Schleswig-Holst
## 7 01054            4155. Nordfriesland    01       Schleswig-Holst
## 8 01055            4824. Ostholstein     01       Schleswig-Holst
## 9 01056            8547. Pinneberg       01       Schleswig-Holst
## 10 01057           2572. 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 × 14
## # Groups:   ost_name [2]
## # ... with 14 variables:
## #   Regionalschluessel     <chr>    total_alo      <dbl>
## #   landkreis_name        <chr>    3970. Flensburg <chr>
## #   ost_name [2]           <chr>    10315. Kiel    01
## #   ... with 9 more variables: Einwohner <dbl>,
## #   bip <dbl>, bip_pro_kopf <dbl>, erw <dbl>,
## #   alo_quote <dbl>, ost <fct>, ost_name <chr>
## #   ... and 390 more rows
## #   ... with 9 more variables: 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 6.30    1.46   3.76    5.16    6.26    7.16   10.7
## 2 Westdeutschland 4.56    1.98   1.50    3.13    3.99    5.74   13.7
```

```
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    6.30    1.46    3.76    5.16    6.26    7.16    10.7  
## 2 Westdeutschland   4.56    1.98    1.50    3.13    3.99    5.74    13.7
```

```
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	6.30	1.46	3.76	5.16	6.26	7.16	10.70
Westdeutschland	4.56	1.98	1.50	3.13	3.99	5.74	13.71

```

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	6.30	1.46	3.76	5.16	6.26	7.16	10.70
Westdeutschland	4.56	1.98	1.50	3.13	3.99	5.74	13.71

```

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	6.30	1.46	3.76	5.16	6.26	7.16	10.70
Westdeutschland	4.56	1.98	1.50	3.13	3.99	5.74	13.71

```

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	6.30	1.46	3.76	5.16	6.26	7.16	10.70
Westdeutschland	4.56	1.98	1.50	3.13	3.99	5.74	13.71

```

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	6.30	1.46	3.76	5.16	6.26	7.16	10.70
Westdeutschland	4.56	1.98	1.50	3.13	3.99	5.74	13.71

Bitte beachten:

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

```

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	6.30	1.46	3.76	5.16	6.26	7.16	10.70
Westdeutschland	4.56	1.98	1.50	3.13	3.99	5.74	13.71

Bitte beachten:

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

Beschreiben Sie die Tabelle!

Die Arbeitslosenquote zwischen Ost- und Westdeutschland

Große Unterschiede werden sichtbar:

- ✚ Mittelwert über 1,7 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 **5,16%**
- ✚ Bei den **westdeutschen Landkreisen** ist das **75% Quantil** bei einer Arbeitslosenquote von **5,74%**!

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_kopf = sd(bip_pro_kopf),
            mean_schulden_kopf = mean(Schulden_gesamt/Einwohner), sd_schulden = sd(Schulden_gesamt/Einwohner))
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(6,9, 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 Kopf",
            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                  3970. Flensburg          01        Schleswig-Holst
## 2 01002                 10315. Kiel               01        Schleswig-Holst
## 3 01003                  8776. Lübeck             01        Schleswig-Holst
## 4 01004                 3359. Neumünster         01        Schleswig-Holst
## 5 01051                 3858. Dithmarschen        01        Schleswig-Holst
## 6 01053                 5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                 4155. Nordfriesland       01        Schleswig-Holst
## 8 01055                 4824. Ostholstein        01        Schleswig-Holst
## 9 01056                 8547. Pinneberg           01        Schleswig-Holst
## 10 01057                2572. 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            3970. Flensburg             01       Schleswig-Holst
## 2 01002            10315. Kiel                  01       Schleswig-Holst
## 3 01003            8776. Lübeck                01       Schleswig-Holst
## 4 01004            3359. Neumünster            01       Schleswig-Holst
## 5 01051            3858. Dithmarschen           01       Schleswig-Holst
## 6 01053            5351. Herzogtum Lauenburg    01       Schleswig-Holst
## 7 01054            4155. Nordfriesland           01       Schleswig-Holst
## 8 01055            4824. Ostholstein             01       Schleswig-Holst
## 9 01056            8547. Pinneberg              01       Schleswig-Holst
## 10 01057           2572. 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 × 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...  3.41    0.741     49825.     15017.     314
## 2 Bayern            2.98    0.703     49282.     21402.     218
## 3 Berlin             NA     NA        NA         NA         NA
## 4 Brandenburg       6.36    1.42      35440.     7563.     246
## 5 Bremen             9.08    2.64      47873.     14123.     111
## 6 Hamburg            NA     NA        NA         NA         NA
## 7 Hessen              4.72    1.17      45974.     18556.     386
## 8 Mecklenburg-Vor...  7.21    1.17      34664.     7288.     386
## 9 Niedersachsen      5.57    1.67      40510.     20508.     254
## 10 Nordrhein-Westf...  6.63    2.37      41435.     12334.     428
## 11 Rheinland-Pfalz  5.05    1.44      40223.     18368.     312
## 12 Saarland           5.57    1.67      36453.     8555.     516
## 13 Sachsen            5.53    0.790     34595.     5879.     241
## 14 Sachsen-Anhalt    7.45    1.38      33644.     5192.     294
## 15 Schleswig-Holst...  5.38    0.865     39300.     9913.     315
## 16 Thüringen          5.64    1.35      32240.     6650.     275
## # 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
## 1	Baden-Württembe...	3.41	0.741	49825.	15017.	314
## 2	Bayern	2.98	0.703	49282.	21402.	218
## 3	Berlin	NA	NA	NA	NA	NA
## 4	Brandenburg	6.36	1.42	35440.	7563.	246
## 5	Bremen	9.08	2.64	47873.	14123.	NA
## 6	Hamburg	NA	NA	NA	NA	NA
## 7	Hessen	4.72	1.17	45974.	18556.	386
## 8	Mecklenburg-Vor...	7.21	1.17	34664.	7288.	386
## 9	Niedersachsen	5.57	1.67	40510.	20508.	254
## 10	Nordrhein-Westf...	6.63	2.37	41435.	12334.	428
## 11	Rheinland-Pfalz	5.05	1.44	40223.	18368.	312
## 12	Saarland	5.57	1.67	36453.	8555.	516
## 13	Sachsen	5.53	0.790	34595.	5879.	241
## 14	Sachsen-Anhalt	7.45	1.38	33644.	5192.	294
## 15	Schleswig-Holst...	5.38	0.865	39300.	9913.	315
## 16	Thüringen	5.64	1.35	32240.	6650.	275
## # 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 × 7
##   bundesland_name  mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_ko
##   <chr>          <dbl>    <dbl>      <dbl>       <dbl>        <dbl>
## 1 Bayern           2.98     0.703     49282.     21402.      218
## 2 Baden-Württembe... 3.41     0.741     49825.     15017.      314
## 3 Hessen            4.72     1.17      45974.     18556.      386
## 4 Rheinland-Pfalz  5.05     1.44      40223.     18368.      312
## 5 Schleswig-Holst... 5.38     0.865     39300.     9913.       315
## 6 Sachsen           5.53     0.790     34595.     5879.       241
## 7 Niedersachsen    5.57     1.67      40510.     20508.      254
## 8 Saarland          5.57     1.67      36453.     8555.       516
## 9 Thüringen         5.64     1.35      32240.     6650.       275
## 10 Brandenburg      6.36     1.42      35440.     7563.       246
## 11 Nordrhein-Westf... 6.63     2.37      41435.     12334.      428
## 12 Mecklenburg-Vor... 7.21     1.17      34664.     7288.       386
## 13 Sachsen-Anhalt   7.45     1.38      33644.     5192.       294
## 14 Bremen            9.08     2.64      47873.     14123.      1
## 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 × 7
##   bundesland_name  mean_alo   sd_alo mean_bip_kopf sd_bip_kopf mean_schulden_ko
##   <chr>          <dbl>    <dbl>      <dbl>       <dbl>        <dbl>
## 1 Bayern           2.98    0.703     49282.     21402.      218
## 2 Baden-Württembe... 3.41    0.741     49825.     15017.      314
## 3 Hessen            4.72    1.17      45974.     18556.      386
## 4 Rheinland-Pfalz  5.05    1.44      40223.     18368.      312
## 5 Schleswig-Holst... 5.38    0.865     39300.     9913.       315
## 6 Sachsen           5.53    0.790     34595.     5879.       241
## 7 Niedersachsen    5.57    1.67      40510.     20508.      254
## 8 Saarland          5.57    1.67      36453.     8555.       516
## 9 Thüringen         5.64    1.35      32240.     6650.       275
## 10 Brandenburg      6.36    1.42      35440.     7563.       246
## 11 Nordrhein-Westf... 6.63    2.37      41435.     12334.      428
## 12 Mecklenburg-Vor... 7.21    1.17      34664.     7288.       386
## 13 Sachsen-Anhalt   7.45    1.38      33644.     5192.       294
## # 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	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

```

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	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

```

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

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

```

Deskriptive Tabelle komplett

Bundesland	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

```

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(6, 9, 10, 12, 13), bold = T, color = "whi

```

Deskriptive Tabelle komplett

Bundesland	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

```

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(6,9, 10, 12,13), bold = T, color = "whi
  add_header_above(c(" " = 1, "Arbeitslosenquote" =

```

Deskriptive Tabelle komplett

Bundesland	Arbeitslosenquote		BIP pro Kopf		Schulden pro Kopf	
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

```

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(6, 9, 10, 12, 13), bold = T, color = "whi
  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	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

Bitte beachten:

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

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

ARBEITSLOSENQUOTE, BIP PRO KOPF UND SCHULDEN PRO KOPF

Deskriptive Tabelle komplett

Bundesland	Arbeitslosenquote		BIP pro Kopf		Schulden pro Kopf	
	Mittelwert	Std.	Mittelwert	Std.	Mittelwert	Std.
Bayern	2.98	0.70	49281.51	21402.37	2181.67	1578.71
Baden-Württemberg	3.41	0.74	49824.77	15016.77	3140.21	2478.42
Hessen	4.72	1.17	45974.34	18556.10	3867.58	3426.79
Rheinland-Pfalz	5.05	1.44	40223.17	18368.50	3124.69	3530.59
Schleswig-Holstein	5.38	0.86	39299.56	9913.17	3158.22	1400.46
Sachsen	5.53	0.79	34595.31	5878.85	2413.17	884.77
Niedersachsen	5.57	1.67	40509.82	20508.37	2540.61	2050.13
Saarland	5.57	1.67	36453.36	8554.62	5162.62	1182.97
Thüringen	5.64	1.35	32239.94	6650.04	2759.01	601.39
Brandenburg	6.36	1.42	35440.47	7562.58	2469.16	1249.58
Nordrhein-Westfalen	6.63	2.37	41434.87	12333.98	4283.95	2447.41
Mecklenburg-Vorpommern	7.21	1.17	34663.84	7287.85	3861.19	1907.93
Sachsen-Anhalt	7.45	1.38	33643.87	5191.52	2947.40	1559.79

Bitte beachten:

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

¹ Die ostdeutschen Bundesländer sind grau hinterlegt.

Was lernen wir aus dieser Tabelle?

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 < 35000€
- ✚ Westdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf > 35000€
- ✚ 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 < 35000€
- + Westdeutsche Landkreise haben im Durchschnitt ein BIP pro Kopf > 35000€
- + 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

Hier gelangen Sie direkt zur [entsprechenden Sektion in der ausformulierten Case-Study](#).

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 2022

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 2022

Die folgende Grafik sollte enthalten:

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

gesamtdaten

```
## # A tibble: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  3970. Flensburg          01        Schleswig-Holst
## 2 01002                  10315. Kiel             01        Schleswig-Holst
## 3 01003                  8776. Lübeck           01        Schleswig-Holst
## 4 01004                  3359. Neumünster       01        Schleswig-Holst
## 5 01051                  3858. Dithmarschen      01        Schleswig-Holst
## 6 01053                  5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4155. Nordfriesland     01        Schleswig-Holst
## 8 01055                  4824. Ostholstein      01        Schleswig-Holst
## 9 01056                  8547. Pinneberg         01        Schleswig-Holst
## 10 01057                 2572. 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,
```

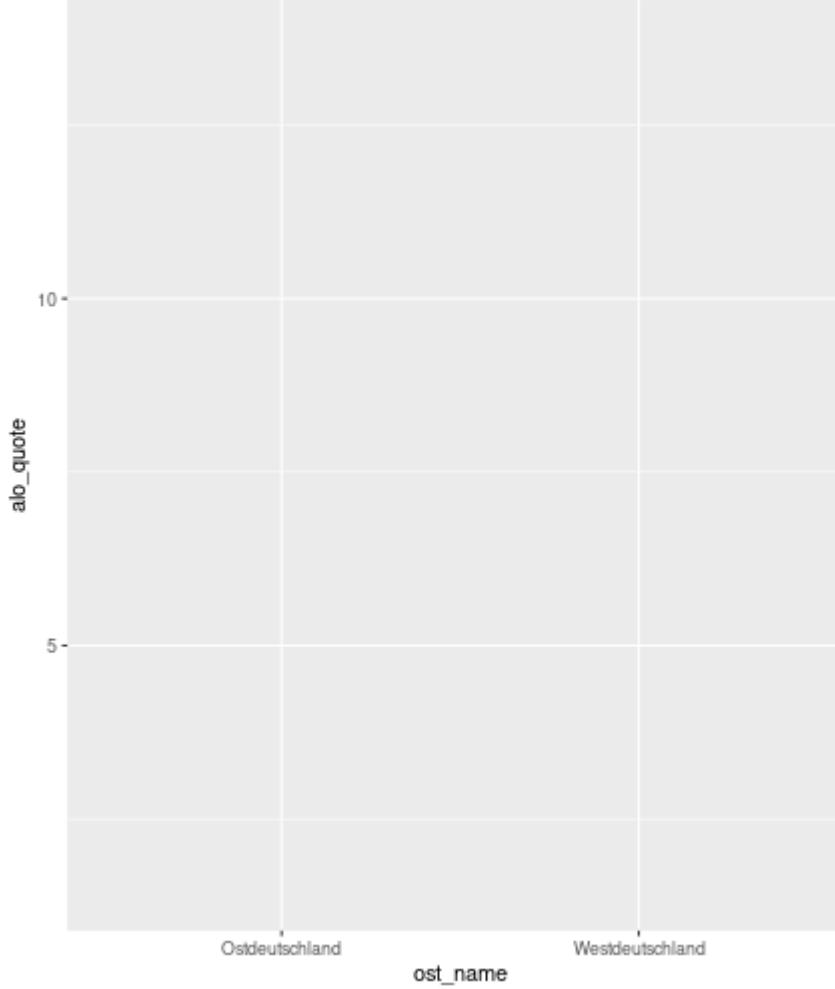
```
## # A tibble: 400 × 4
##   alo_quote landkreis_name  bundesland_name  ost_name
##       <dbl> <chr>          <chr>           <chr>
## 1     6.05 Flensburg    Schleswig-Holstein Westdeutschland
## 2     5.46 Kiel          Schleswig-Holstein Westdeutschland
## 3     6.07 Lübeck        Schleswig-Holstein Westdeutschland
## 4     5.68 Neumünster   Schleswig-Holstein Westdeutschland
## 5     5.80 Dithmarschen Schleswig-Holstein Westdeutschland
## 6     6.78 Herzogtum Lauenburg Schleswig-Holstein Westdeutschland
## 7     4.21 Nordfriesland Schleswig-Holstein Westdeutschland
## 8     4.82 Ostholstein   Schleswig-Holstein Westdeutschland
## 9     6.07 Pinneberg    Schleswig-Holstein Westdeutschland
## 10    5.32 Plön          Schleswig-Holstein Westdeutschland
## # i 390 more rows
```

```
gesamtdaten %>%
```

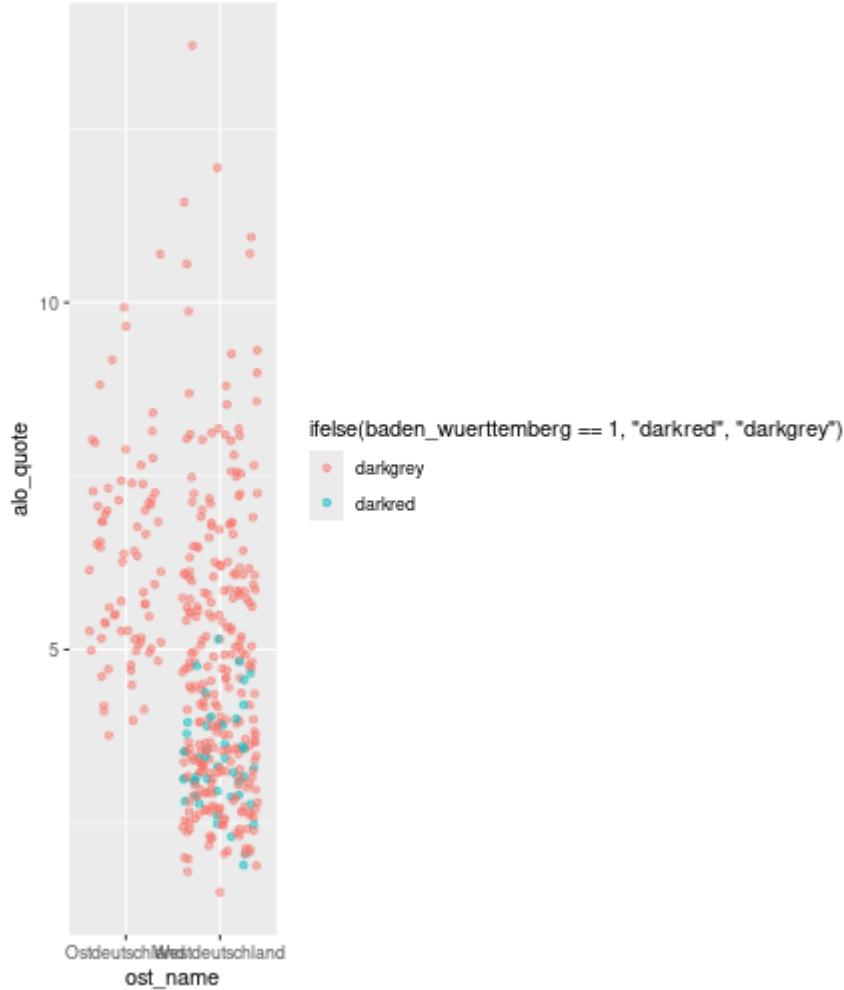
```
  select(alo_quote, landkreis_name, bundesland_name,  
         mutate(baden_wuerttemberg = as.factor(ifelse(bunde
```

```
## # A tibble: 400 × 5  
##   alo_quote landkreis_name   bundesland_name   ost_name baden_wuerttembe  
##       <dbl> <chr>           <chr>            <chr>    <fct>  
## 1     6.05 Flensburg      Schleswig-Holstein Westdeut... 0  
## 2     5.46 Kiel           Schleswig-Holstein Westdeut... 0  
## 3     6.07 Lübeck         Schleswig-Holstein Westdeut... 0  
## 4     5.68 Neumünster    Schleswig-Holstein Westdeut... 0  
## 5     5.80 Dithmarschen   Schleswig-Holstein Westdeut... 0  
## 6     6.78 Herzogtum Lauenburg Schleswig-Holstein Westdeut... 0  
## 7     4.21 Nordfriesland  Schleswig-Holstein Westdeut... 0  
## 8     4.82 Ostholstein    Schleswig-Holstein Westdeut... 0  
## 9     6.07 Pinneberg     Schleswig-Holstein Westdeut... 0  
## 10    5.32 Plön          Schleswig-Holstein Westdeut... 0  
## # i 390 more rows
```

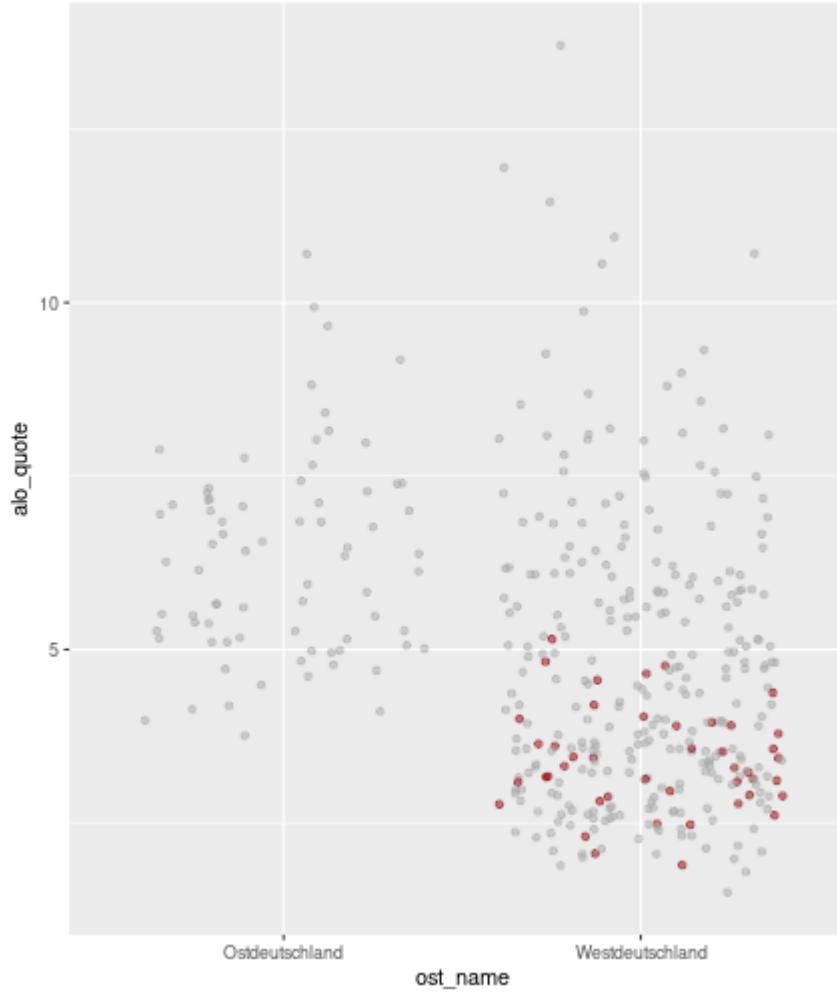
```
gesamtdataen %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote))
```



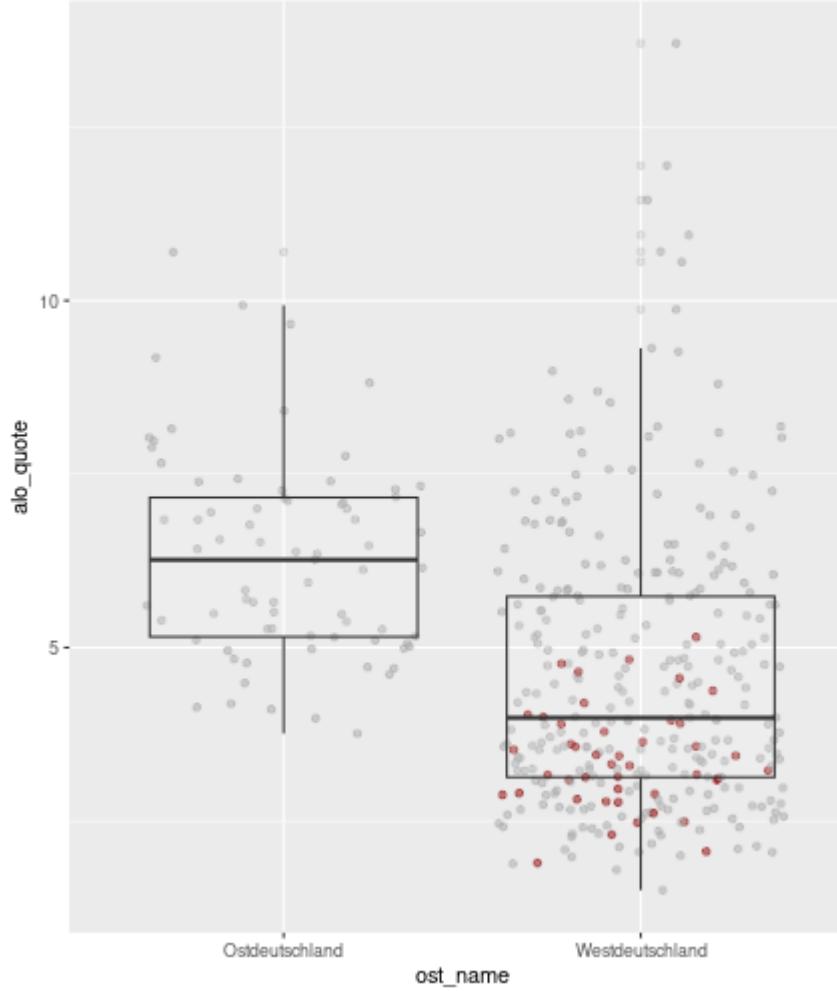
```
gesamtdaten %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
```



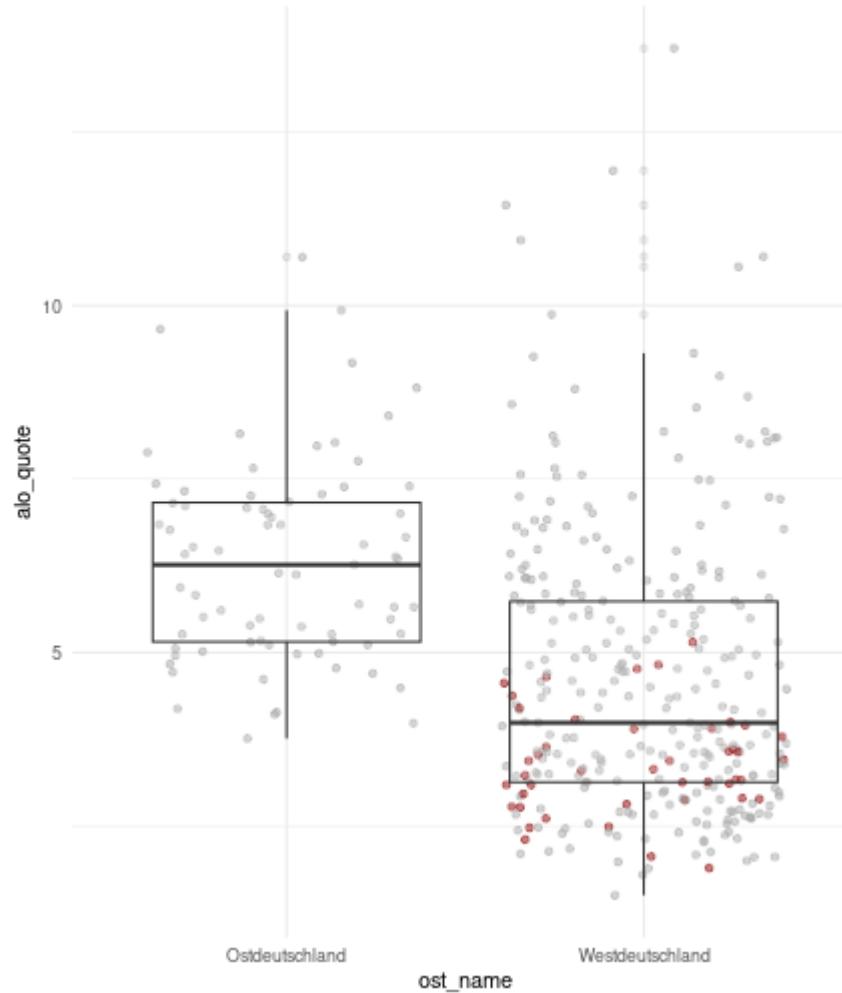
```
gesamtdataen %>%
  select(alo_quote, landkreis_name, bundesland_name,
  mutate(baden_wuerttemberg = as.factor(ifelse(bunde
ggplot(aes(x = ost_name, y=alo_quote)) +
  geom_jitter(aes(color = ifelse(baden_wuerttemberg
  scale_color_identity()
```



```
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()
```



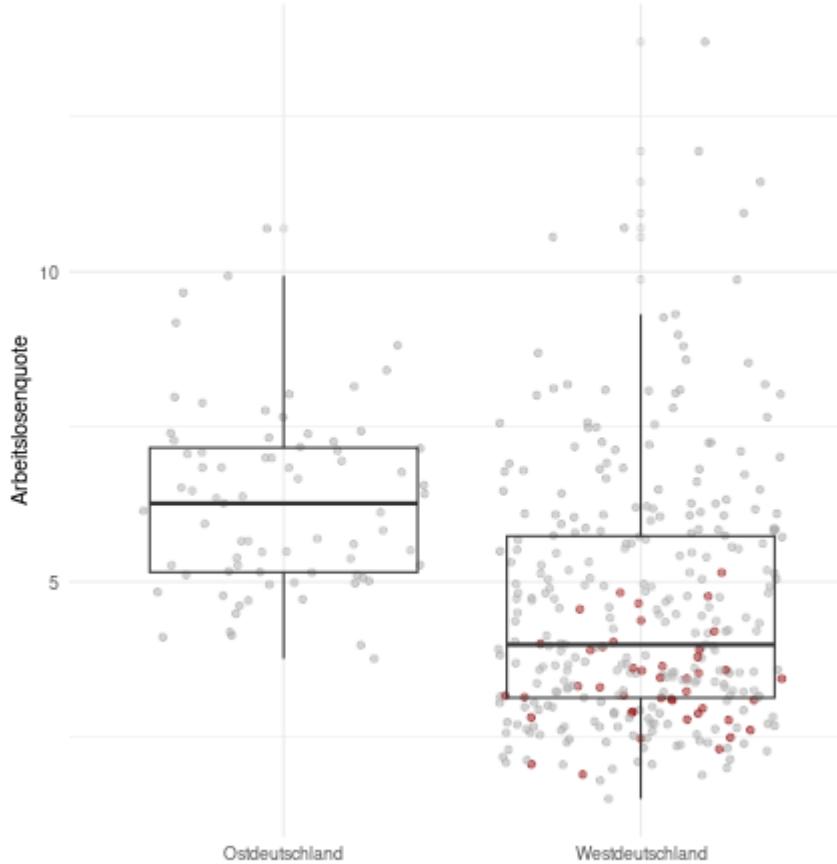
```

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

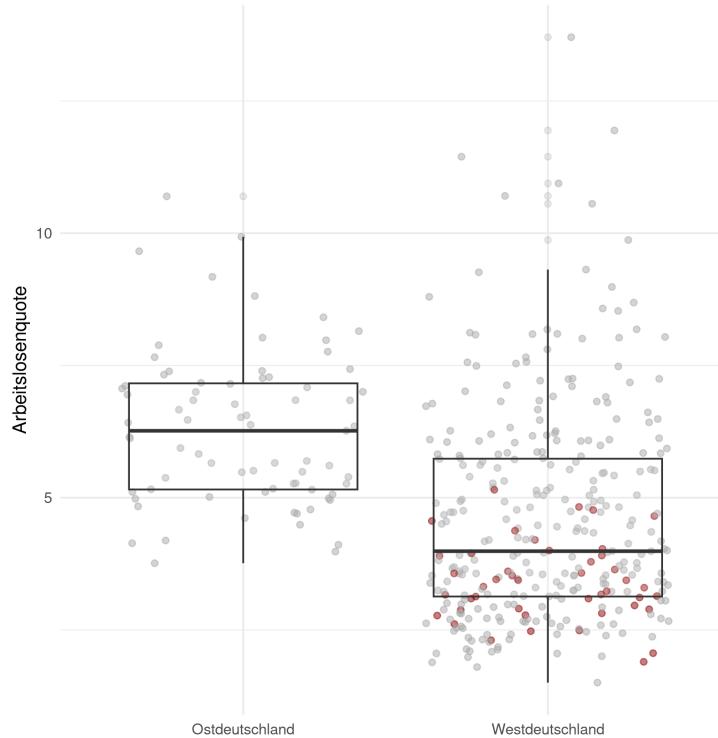


Quelle: Daten der Agentur für Arbeit aus dem Jahr 2022

Arbeitslosenquote

Beschreiben Sie das gezeigte Schaubild

Arbeitslosenquote in Deutschland
Eine Beobachtung repräsentiert einen Landkreis, Baden-Württemberg rot eingefärbt



Quelle: Daten der Agentur für Arbeit aus dem Jahr 2022

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 deutlich unter dem Median 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.

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

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

```
## # A tibble: 11,940 × 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 11,930 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

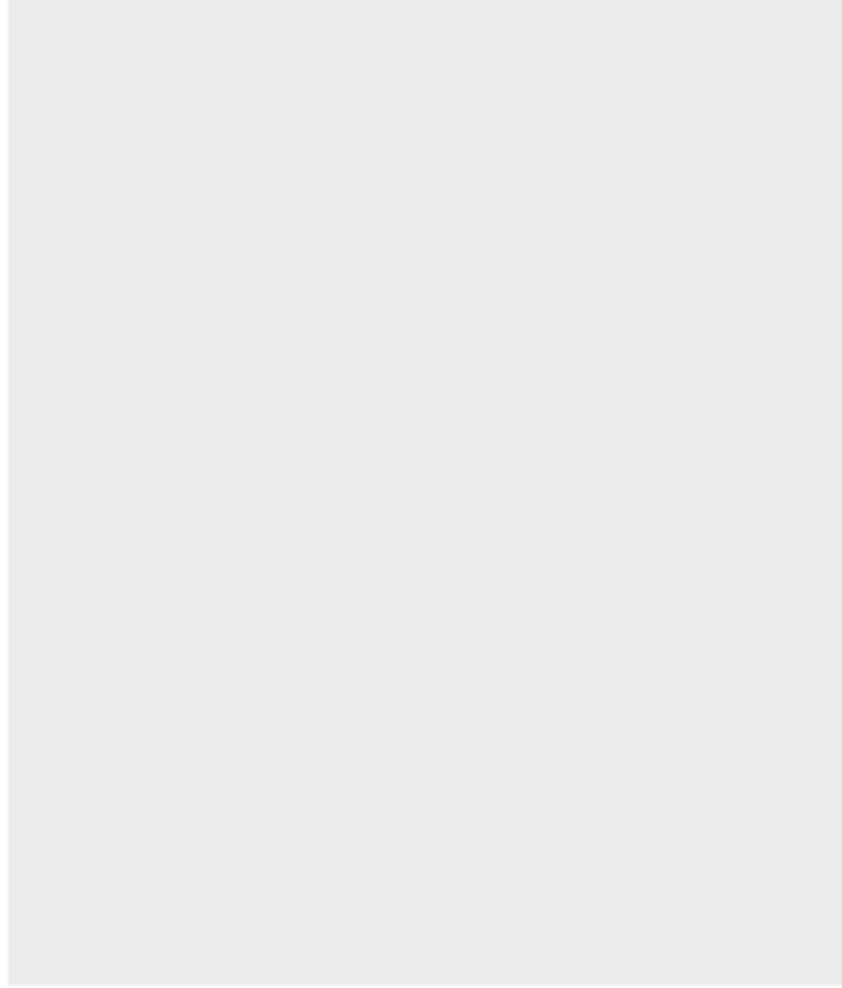
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 )
## # A tibble: 9,154 × 8
##   Regionalschluessel Jahr      bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>     <dbl>     <dbl>     <dbl> <chr>
## 1 08111            2000 35273886000  571528  61719. Baden-Württembe
## 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 9,144 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)
```

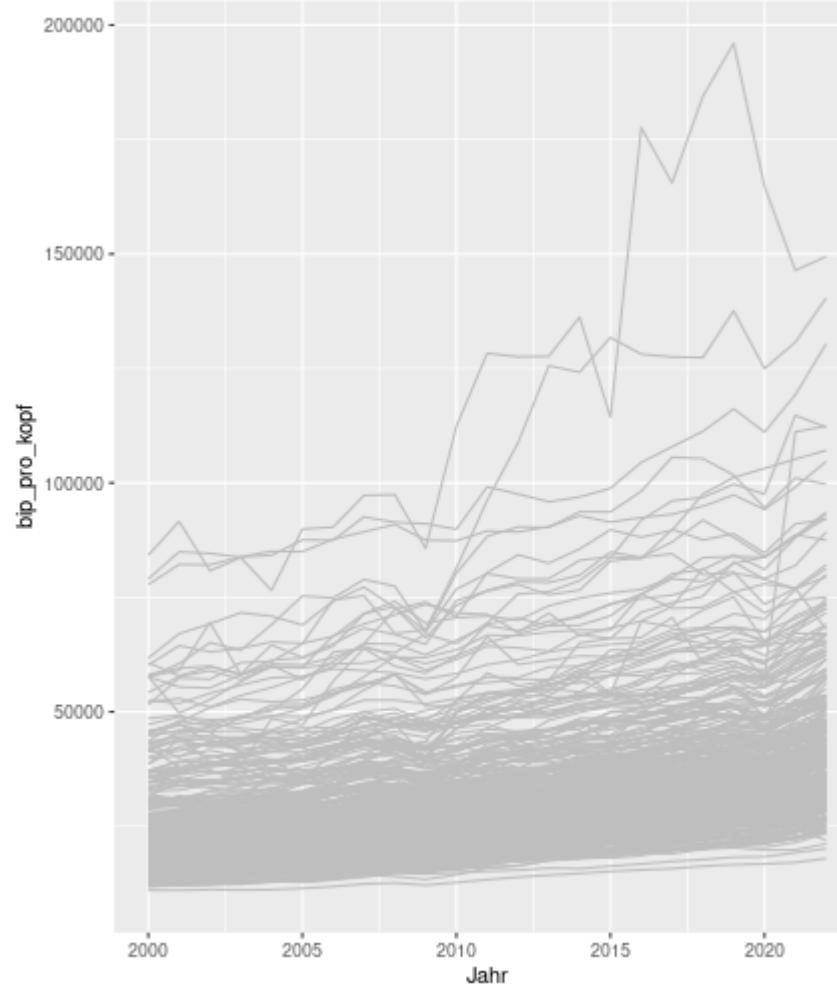
```
## # A tibble: 9,154 × 8
## # Groups:   ost_name, Jahr [46]
##   Regionalschluessel Jahr      bip einwohner bip_pro_kopf bundesland_name
##   <chr>           <dbl>    <dbl>    <dbl>    <dbl>    <chr>
## 1 08111          2000 35273886000  571528  61719. Baden-Württembe
## 2 08115          2000 13867882000  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 9,144 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: 9,154 × 10
#> # Groups:   ost_name, Jahr [46]
#> #   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 9,144 more rows
#> # # i 4 more variables: landkreis_name <chr>, ost_name <chr>, durchschnitt <dbl>
#> #     ulm <dbl>
```

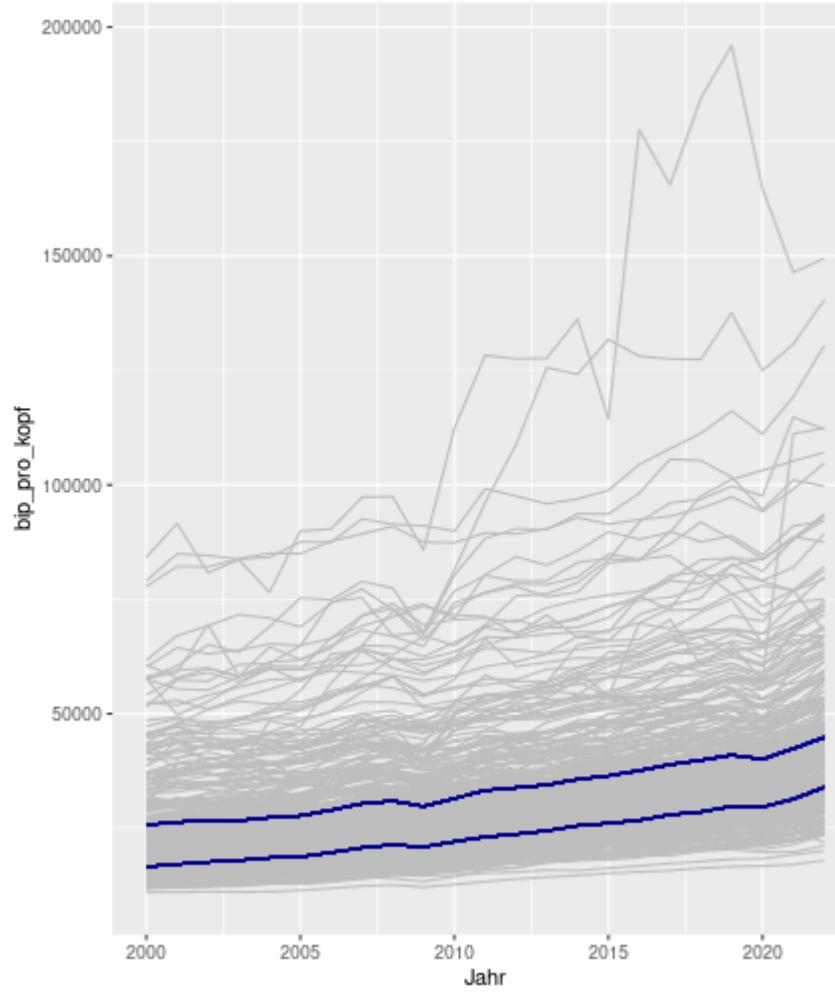
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_
ggplot()
```



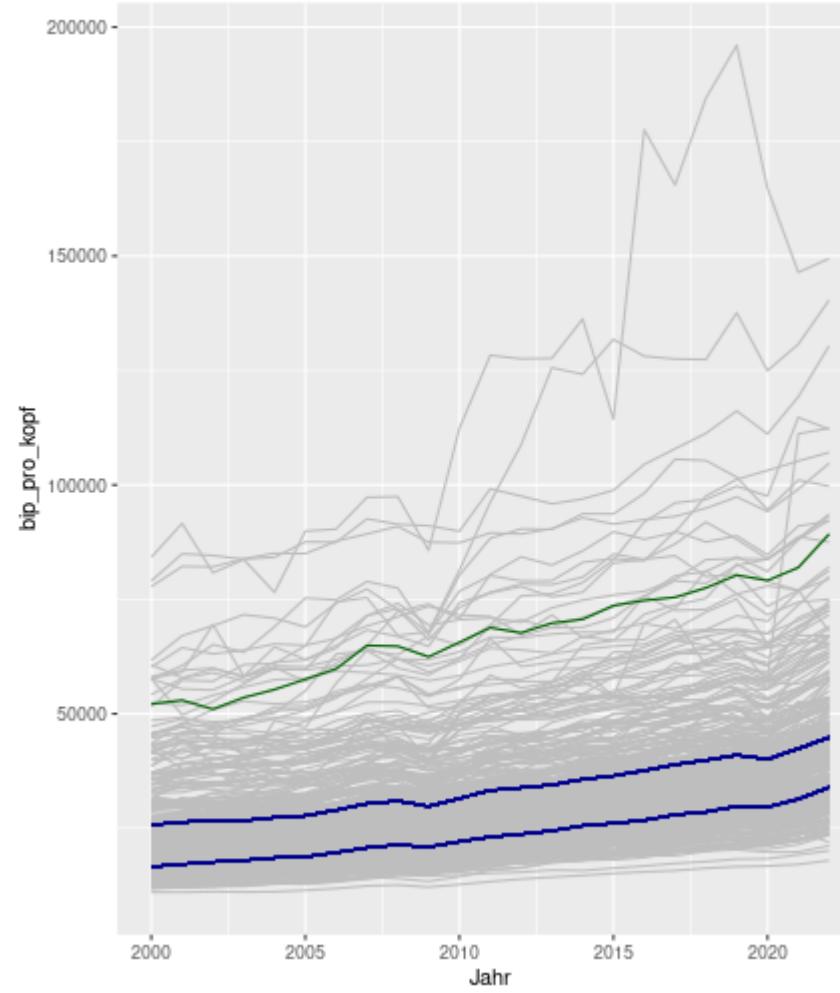
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_-
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
```



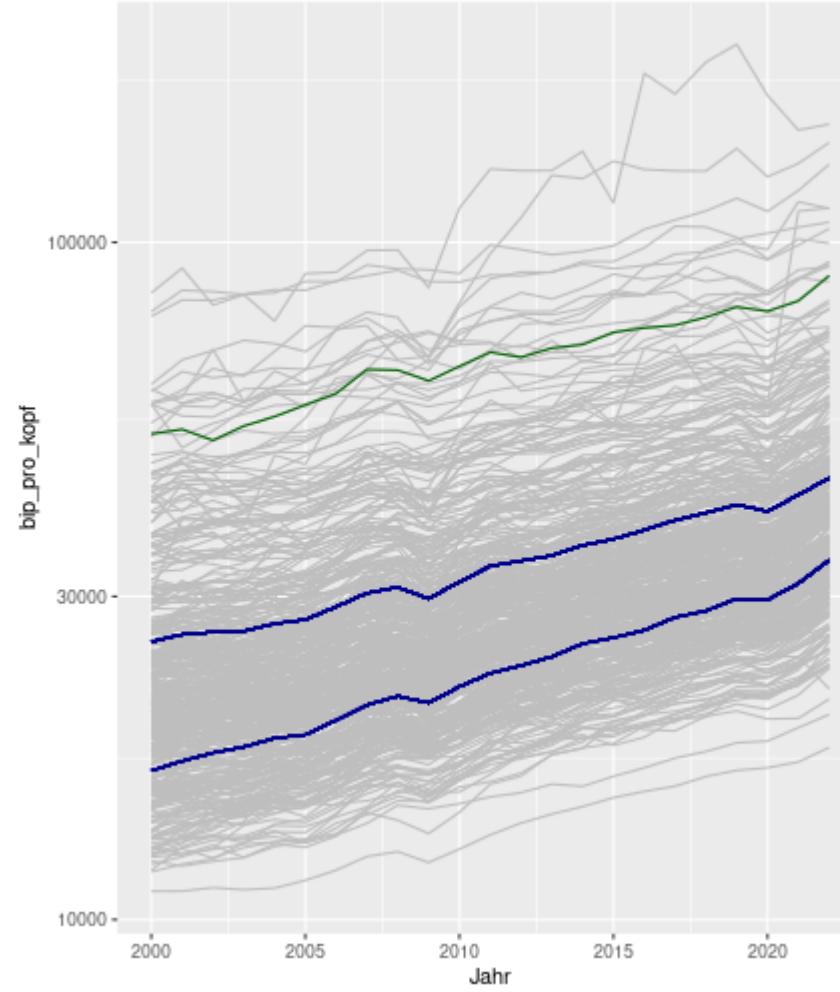
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt,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



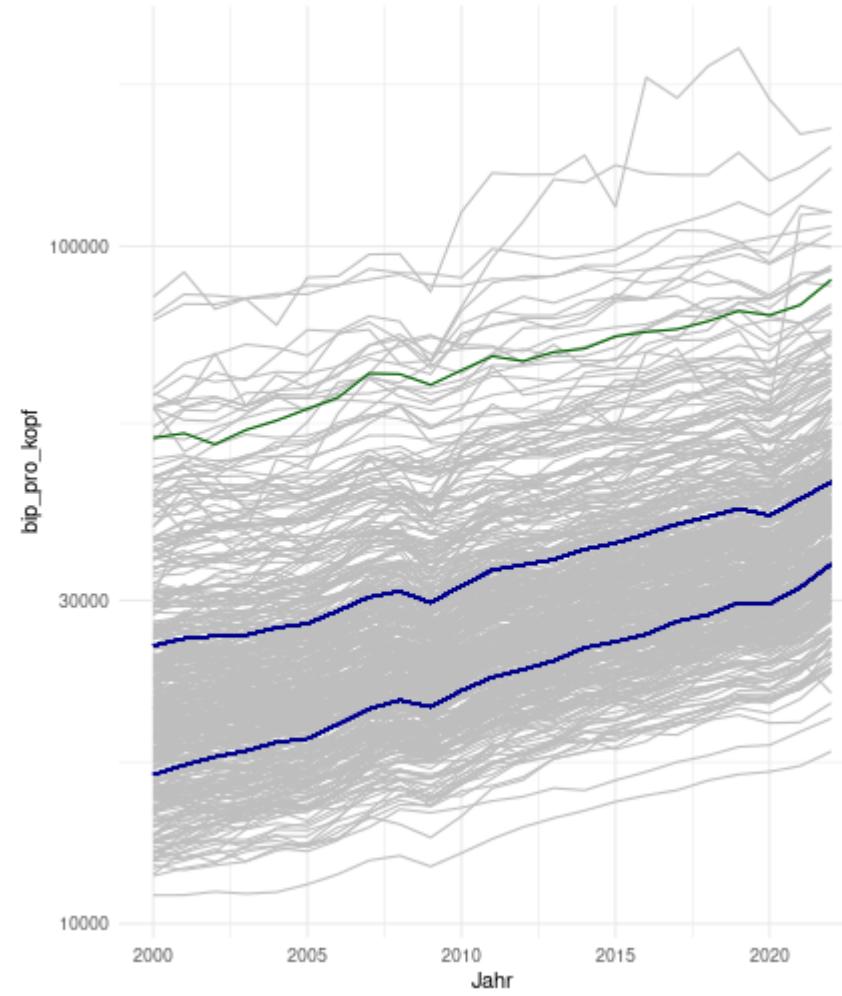
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_
ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr, y = durchschnitt,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



```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf, 0),
         Regionals =
           geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
             geom_line(aes(x = Jahr, y = durchschnitt, group =
               geom_line(aes(x = Jahr, y = ulm, group = Regionals
scale_y_continuous(trans = "log10")
```



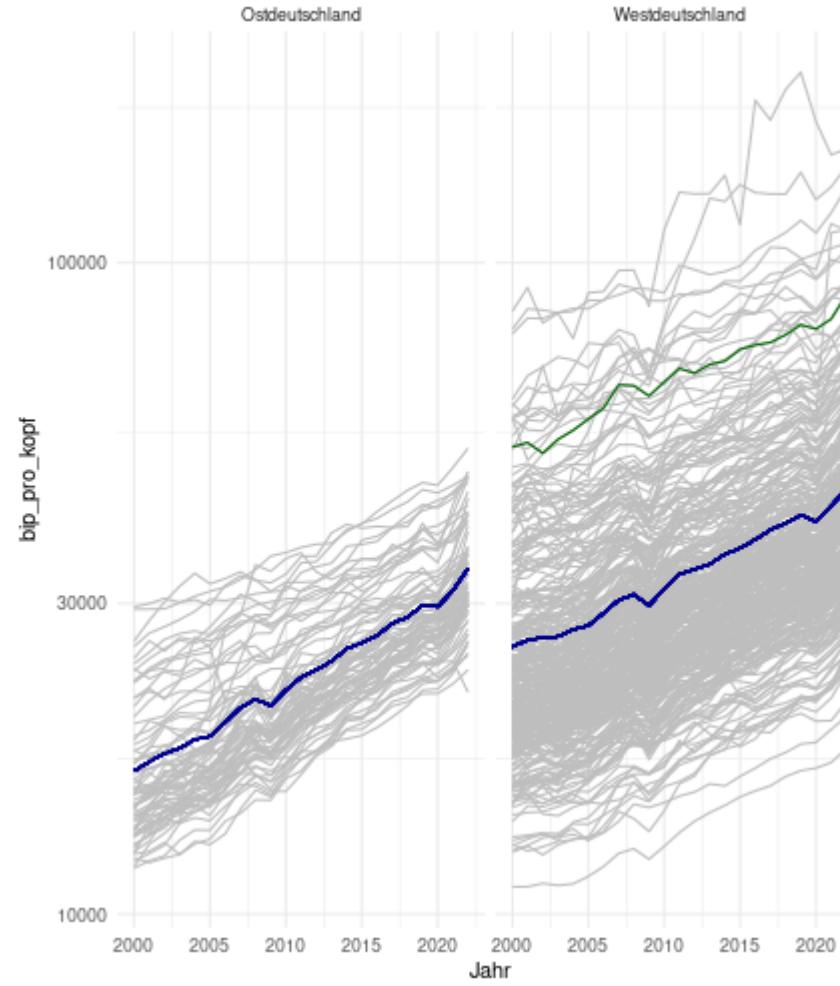
```
options(scipen = 5)
bip_zeitreihe_namen %>%
  filter( Jahr >= 2000 ) %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf),
         ulm = ifelse(landkreis_name == "Ulm", bip_pro_kopf,
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf, group =
  geom_line(aes(x = Jahr,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



```

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

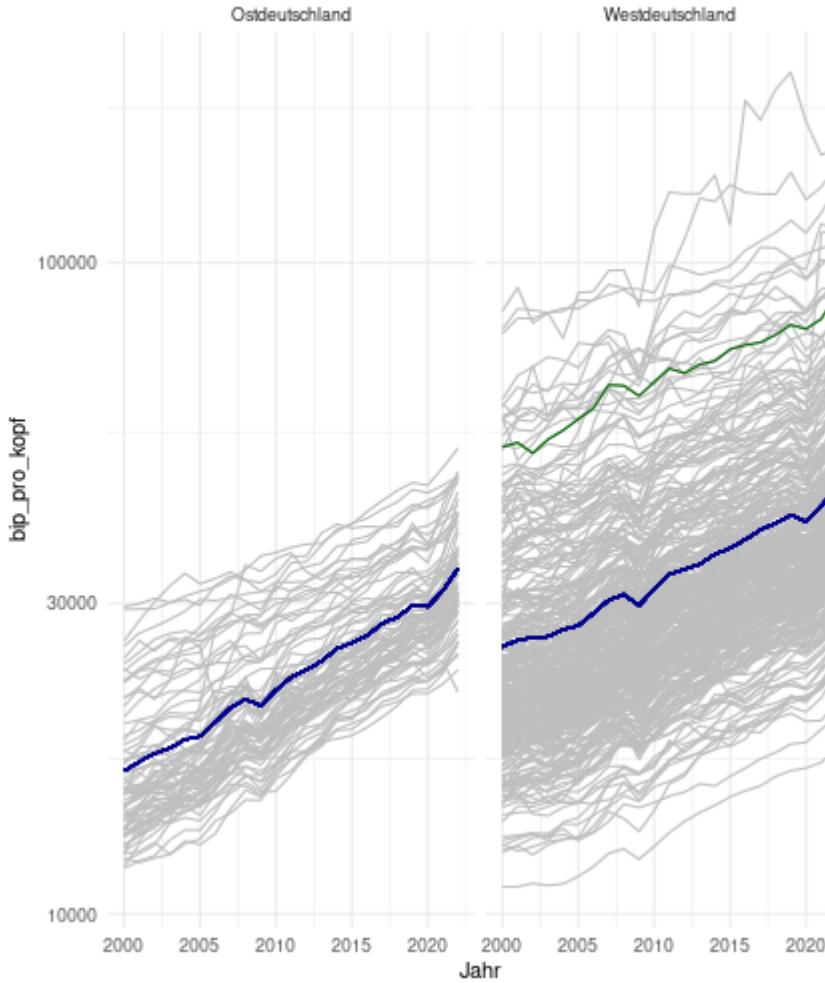
```



```

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

```

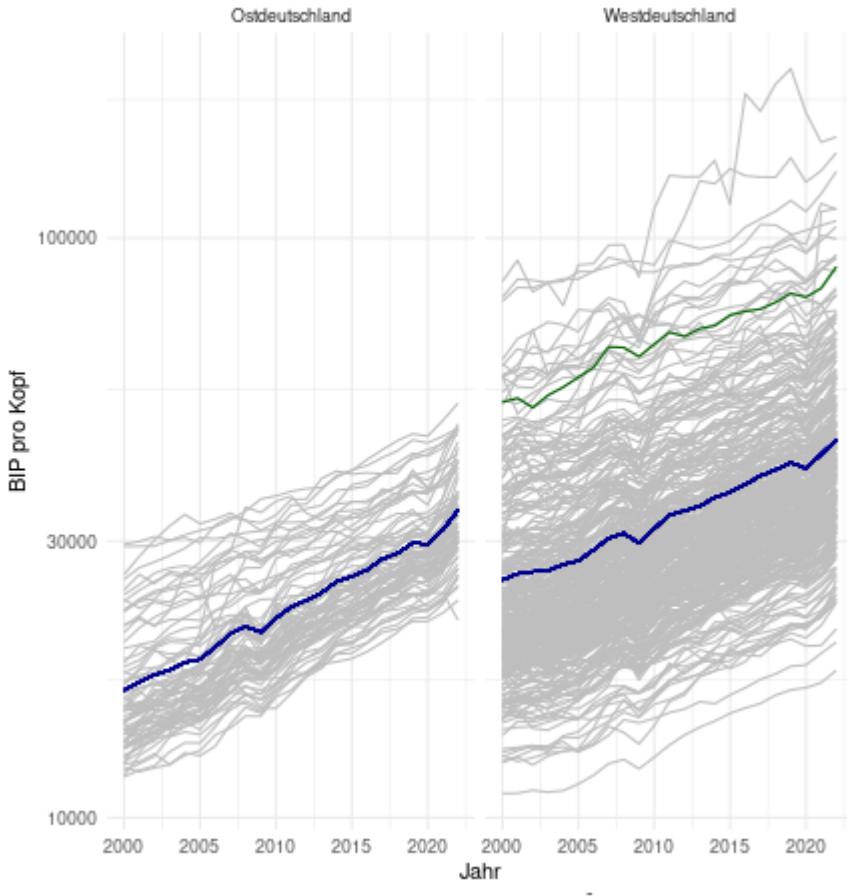


```

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



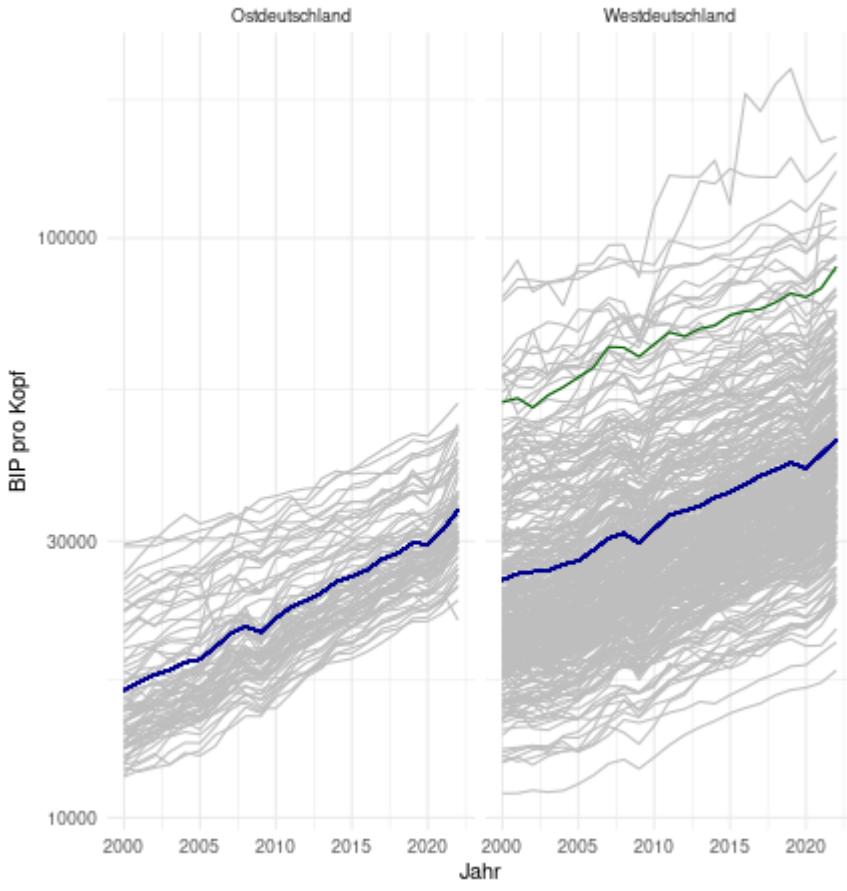
Quelle: Daten der Statistischen Ämter der Länder und des Bundes.

```

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



Beschreiben und interpretieren Sie das gezeigte Schaubild.

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 33936€ den Wert, welchen die westdeutschen Landkreise durchschnittlich in 2012 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 33936€ den Wert, welchen die westdeutschen Landkreise durchschnittlich in 2012 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:

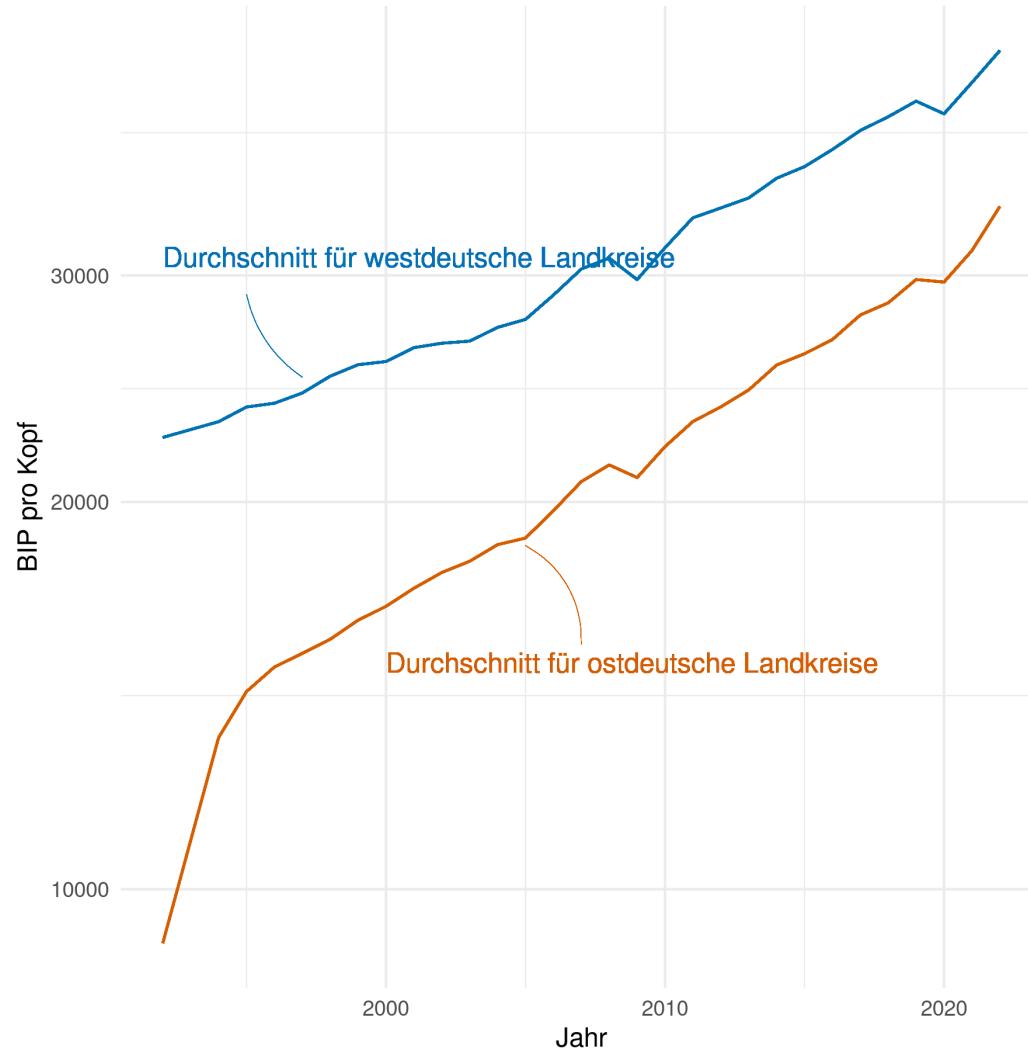
- + Eine Wachstumsprozess im BIP pro Kopf findet in allen Landkreisen statt, jedoch gibt es für die ostdeutschen Landkreise, welche deutlich niedriger gestartet sind, keinen erkennbaren Anpassungsprozess in Form eines schnelleren Wachstums
- + Wir sehen auch keinen Anpassungsprozess der Landkreise in 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 2022



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 2022

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 2022

berechnen und visualisieren.

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

```
bip_zeitreihe_namen
```

```
## # A tibble: 11,940 × 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 11,930 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

```
## # A tibble: 11,940 × 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 11,930 more rows
## # i 2 more variables: landkreis_name <chr>, ost_name <chr>
```

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

## # A tibble: 11,940 × 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 11,930 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: 11,940 × 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 11,930 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: 11,940 × 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 11,930 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: 11,940 × 9
## # Groups:   ost_name, Jahr [60]
##   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 11,930 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: 11,940 × 10
## # Groups:   ost_name, Jahr [60]
##   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 11,930 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: 11,940 × 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 11,930 more rows
## # i 4 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>, durchschnitt <dbl>
```

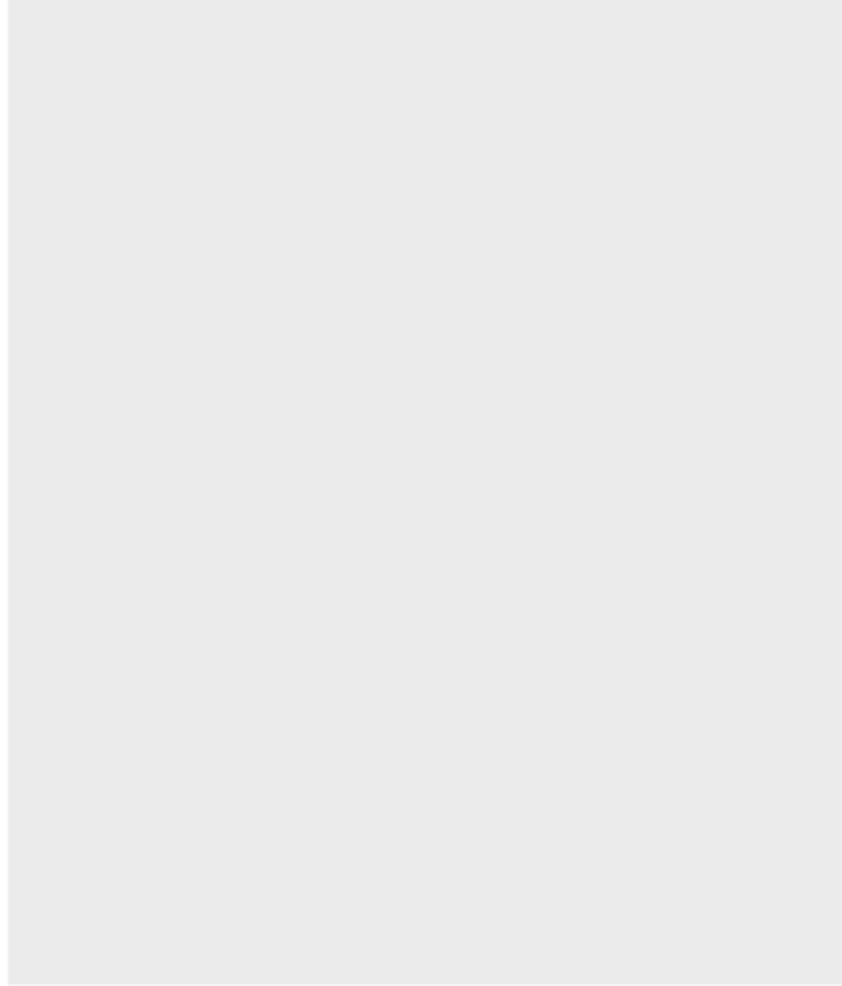
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf -
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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

```
## # A tibble: 9,154 × 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 9,144 more rows
## # i 4 more variables: landkreis_name <chr>, ost_name <chr>,
## #   bip_pro_kopf_wachstum <dbl>, durchschnitt <dbl>
```

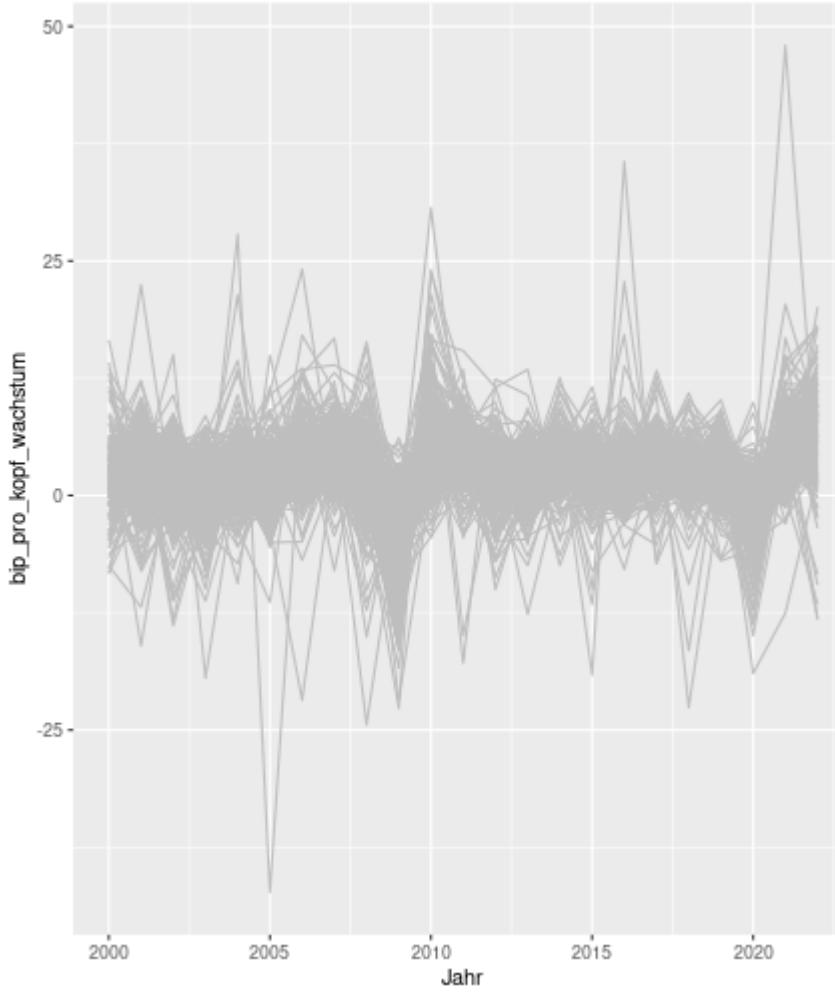
```
bip_zeitreihe_namen %>%
  group_by(Regionalschluessel) %>%
  arrange(Regionalschluessel, Jahr) %>%
  mutate( bip_pro_kopf_wachstum = 100*(bip_pro_kopf
  ungroup() %>%
  group_by(ost_name, Jahr) %>%
  mutate( durchschnitt = mean(bip_pro_kopf_wachstum,
  ungroup() -> bip_wachstum

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,
```

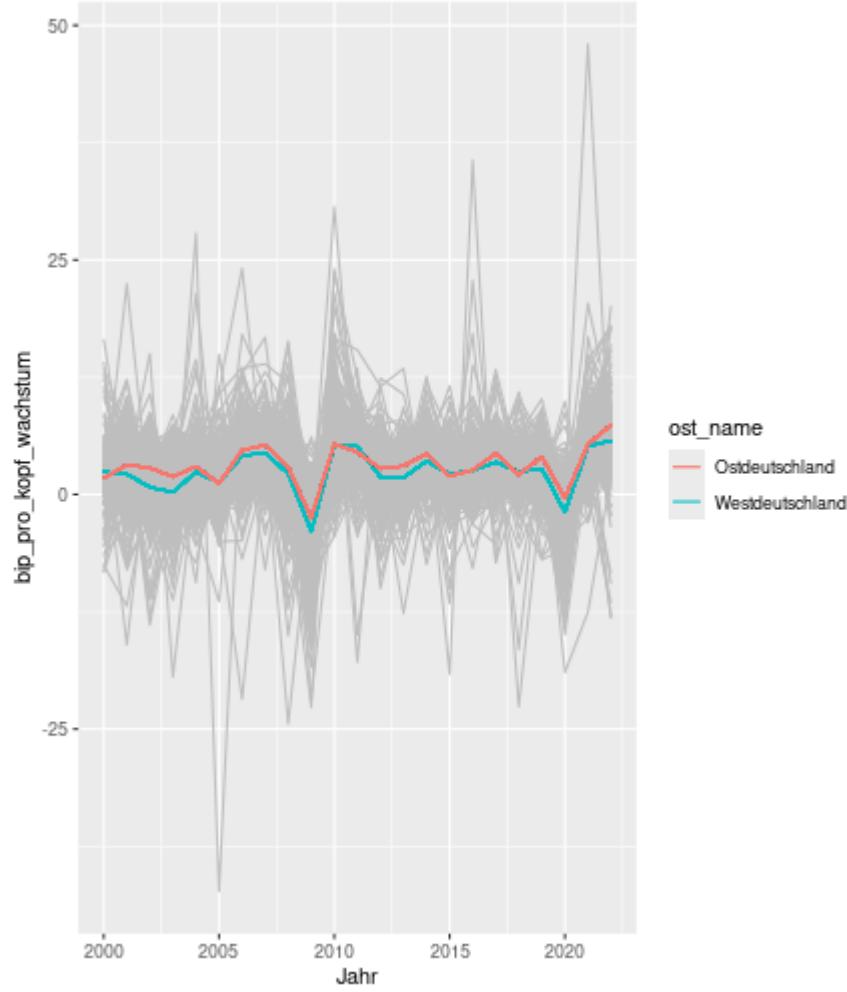


```

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

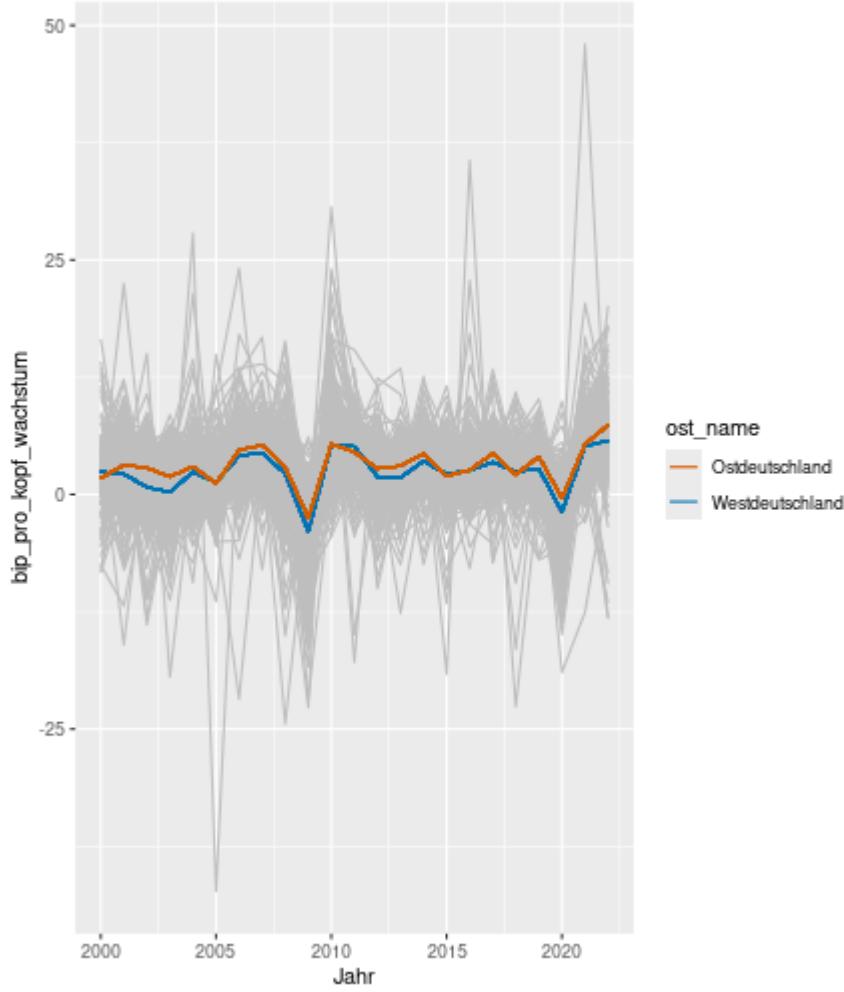
bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf_wachstum,
  geom_line(aes(x = Jahr, y = durchschnitt, group =

```



```
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"
```

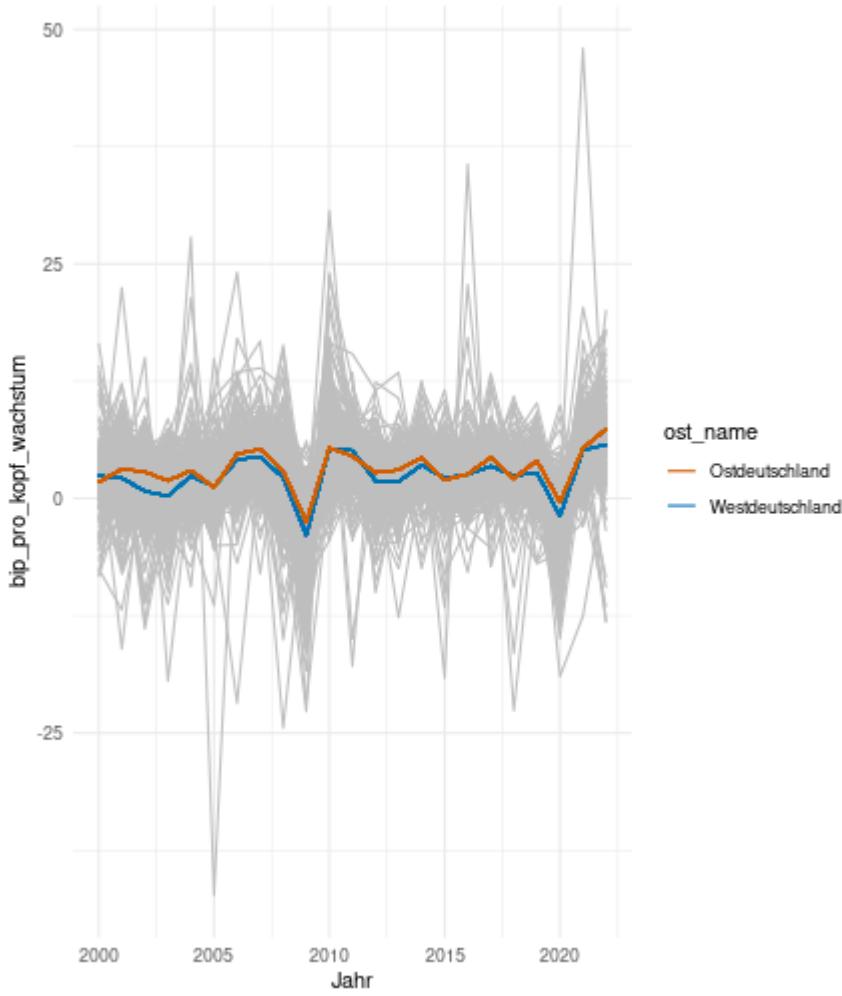


```

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()

```



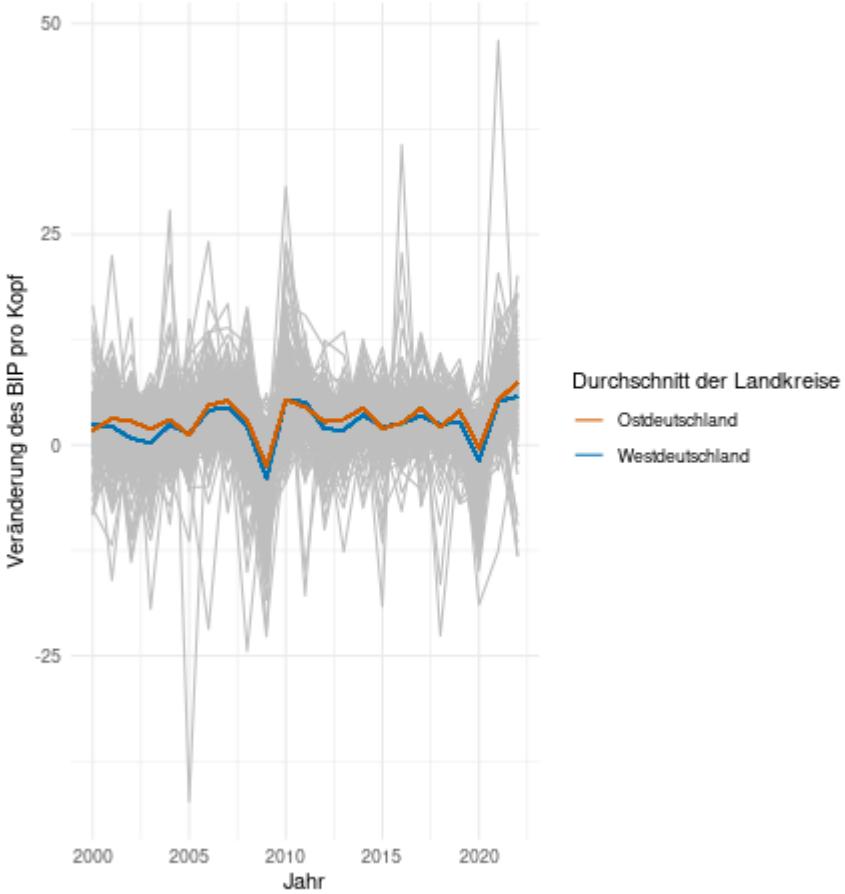
```

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

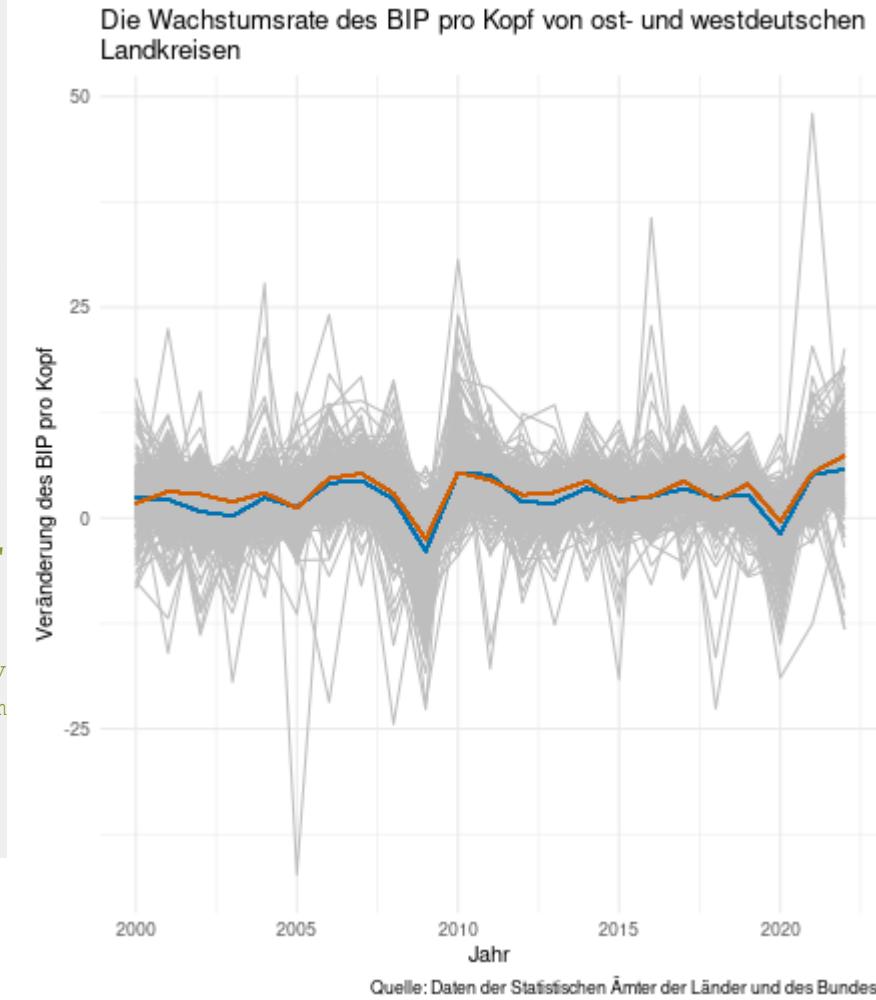


```

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")

```



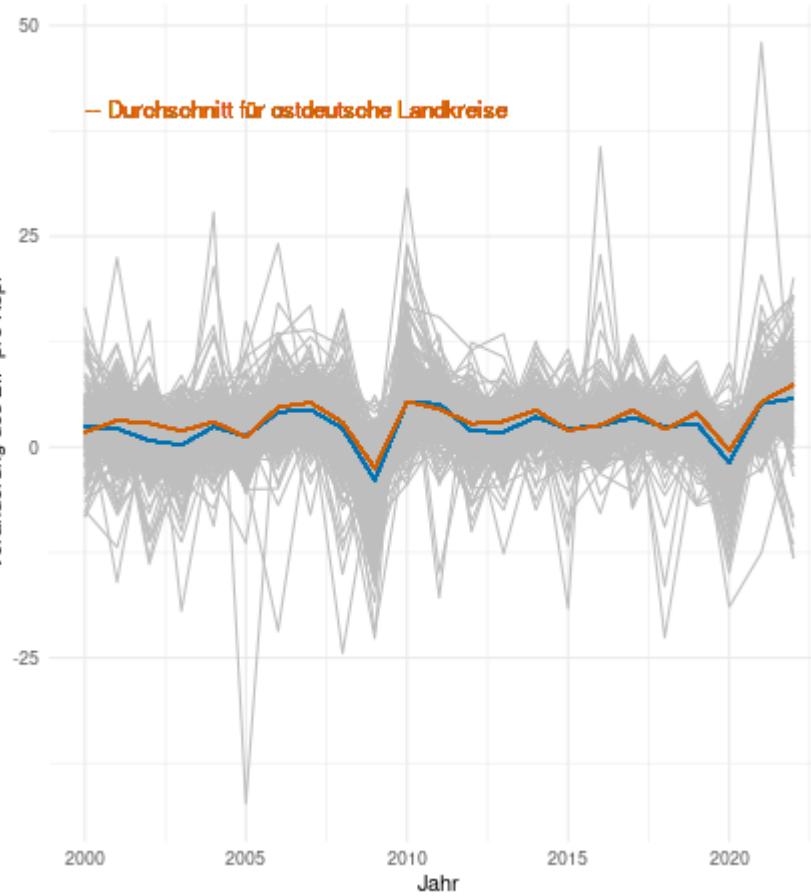
```

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



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

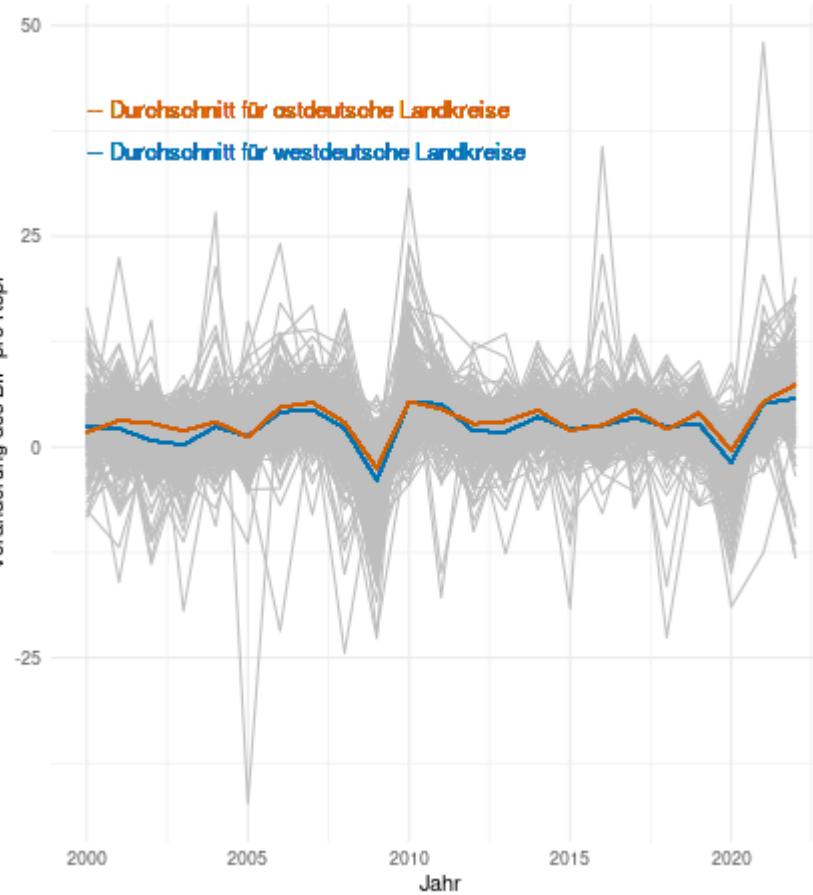
```

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

bip_wachstum %>%
  filter( Jahr >= 2000 ) %>%
  ggplot() +
  geom_line(aes(x = Jahr, y = bip_pro_kopf_wachstum,
  geom_line(aes(x = Jahr, y = durchschnitt, group =
  scale_color_manual(values = c("#D55E00", "#0072B2"
  theme_minimal() +
  labs(color = "Durchschnitt der Landkreise",
  title = "Die Wachstumsrate des BIP pro Kopf von ost- und westdeutschen
  caption = "Quelle: Daten der Statistischen Ämter der Länder und des Bundes."
  x = "Jahr",
  y = "Veränderung des BIP pro Kopf") +
  theme(legend.position = "none") +
  geom_text(aes(x=2000, y=40, label = "-- Durchschnit
  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](#))
- + Viele unterschiedliche Grafiken in der [R Graph Gallery](#) mit dem dazugehörigen Code

gesamtdaten

```
## # A tibble: 400 × 14
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  3970. Flensburg          01        Schleswig-Holst
## 2 01002                  10315. Kiel             01        Schleswig-Holst
## 3 01003                  8776. Lübeck           01        Schleswig-Holst
## 4 01004                  3359. Neumünster       01        Schleswig-Holst
## 5 01051                  3858. Dithmarschen      01        Schleswig-Holst
## 6 01053                  5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4155. Nordfriesland     01        Schleswig-Holst
## 8 01055                  4824. Ostholstein      01        Schleswig-Holst
## 9 01056                  8547. Pinneberg         01        Schleswig-Holst
## 10 01057                 2572. 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)

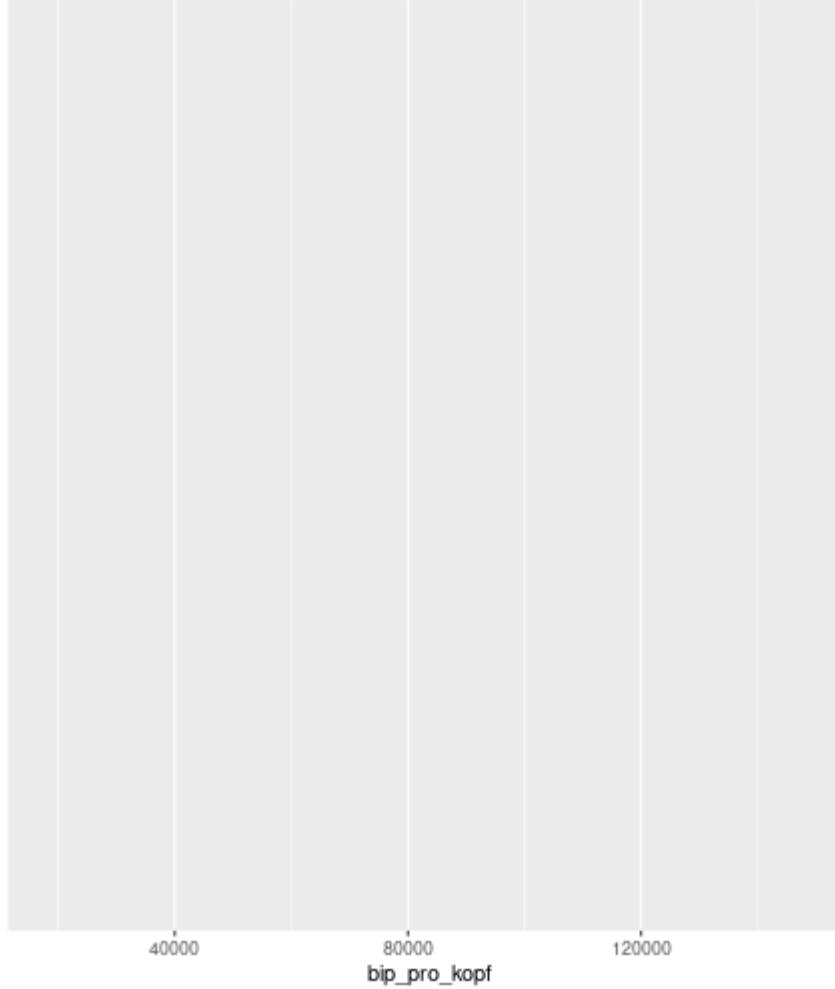
## # A tibble: 400 × 14
## # Groups: ost_name [2]
##   Regionalschlüssel total_alo landkreis_name      bundesland bundesland_name
##   <chr>           <dbl> <chr>                <chr>      <chr>
## 1 01001            3970. Flensburg             01        Schleswig-Holst
## 2 01002            10315. Kiel                  01        Schleswig-Holst
## 3 01003            8776. Lübeck                01        Schleswig-Holst
## 4 01004            3359. Neumünster            01        Schleswig-Holst
## 5 01051            3858. Dithmarschen          01        Schleswig-Holst
## 6 01053            5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054            4155. Nordfriesland          01        Schleswig-Holst
## 8 01055            4824. Ostholstein            01        Schleswig-Holst
## 9 01056            8547. Pinneberg              01        Schleswig-Holst
## 10 01057           2572. 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) %>%  
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm
```

```
## # A tibble: 2 × 2  
##   ost_name      durchschnitt  
##   <chr>          <dbl>  
## 1 Ostdeutschland     33937.  
## 2 Westdeutschland    44860.
```

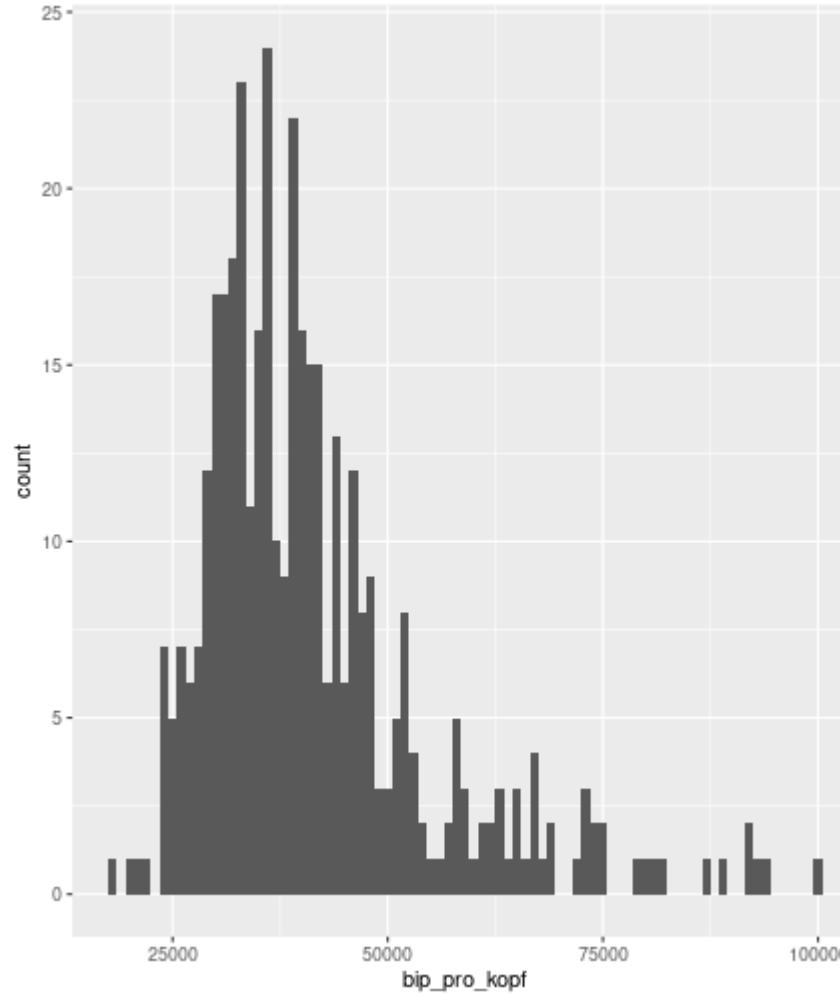
```
gesamtdataen %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel
```

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



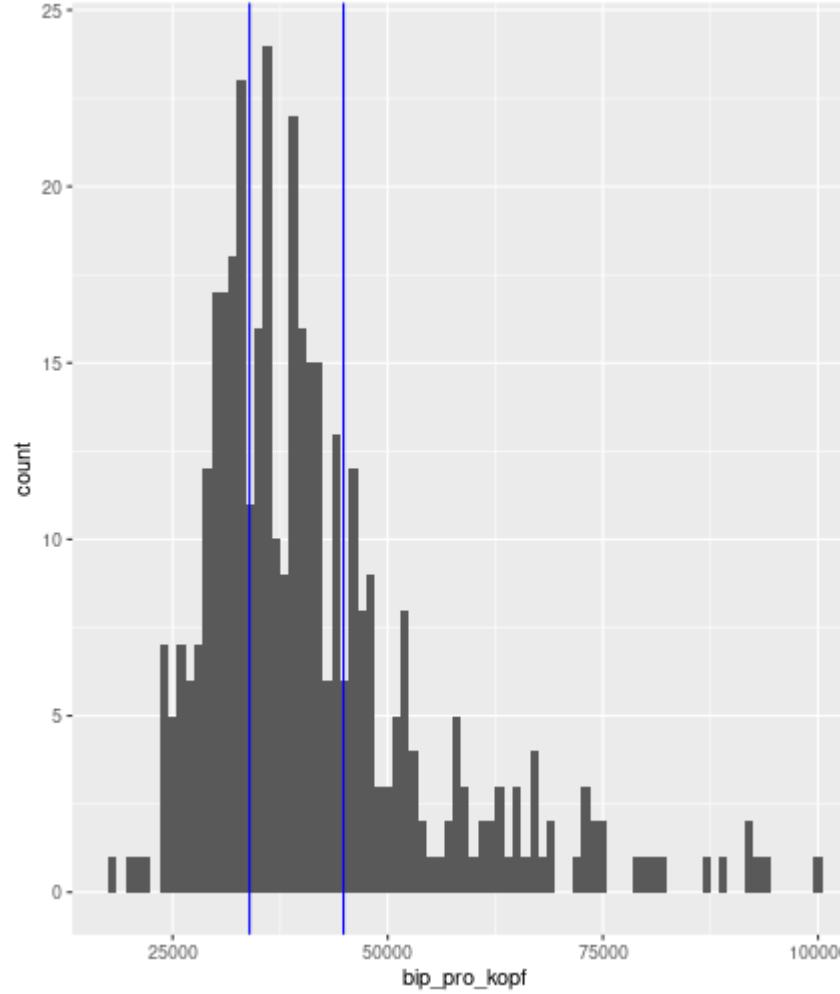
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf >= 25000),
```



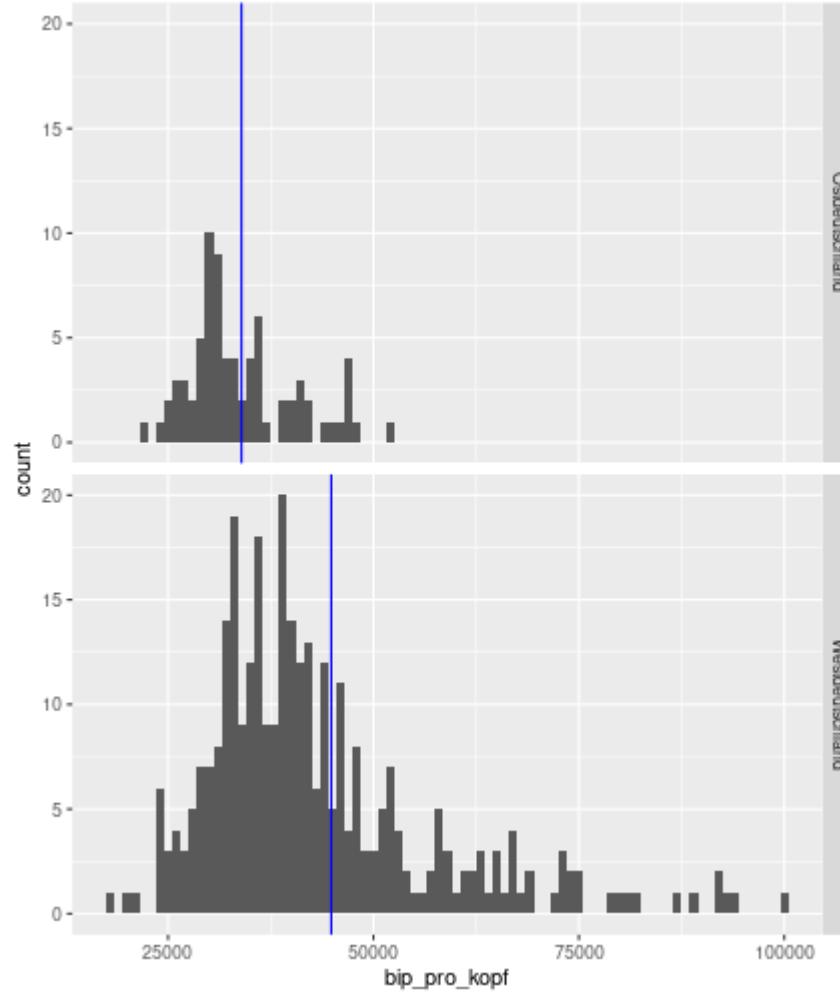
```
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf >= 25000), fill = "black", 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 <= 100000), bins = 100)
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
  facet_grid(ost_name ~ .)
```

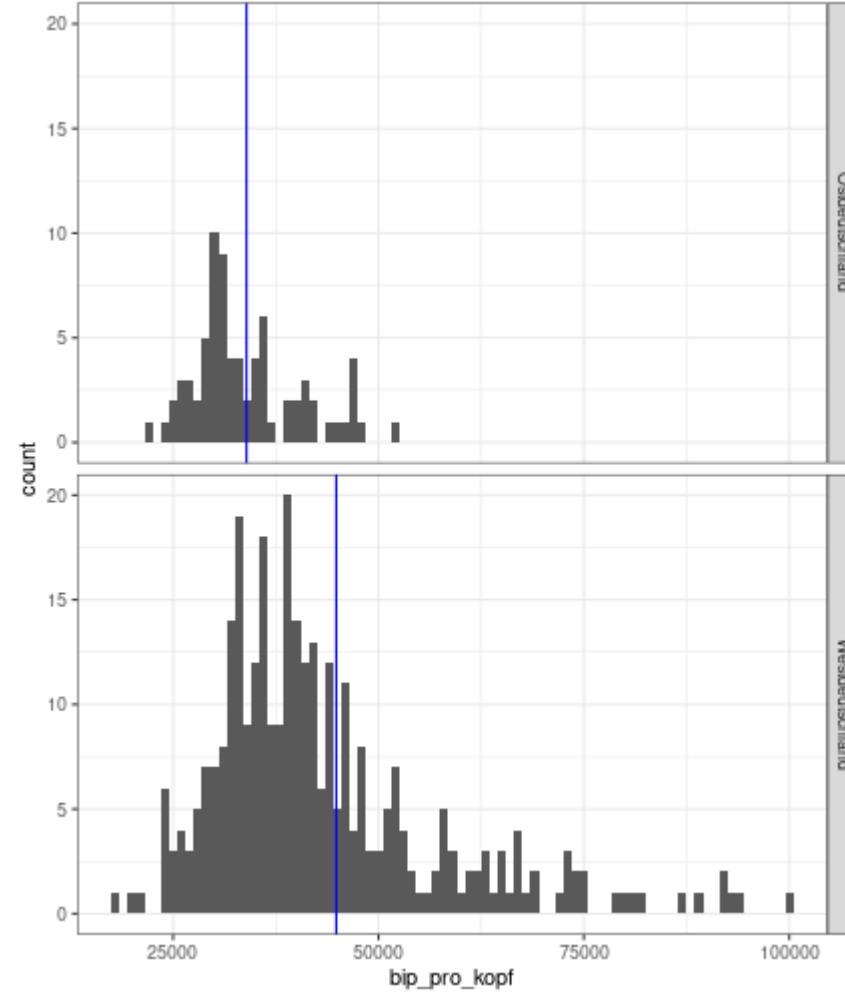


```

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), fill = "black") +
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
facet_grid(ost_name ~ .) +
  theme_bw()

```

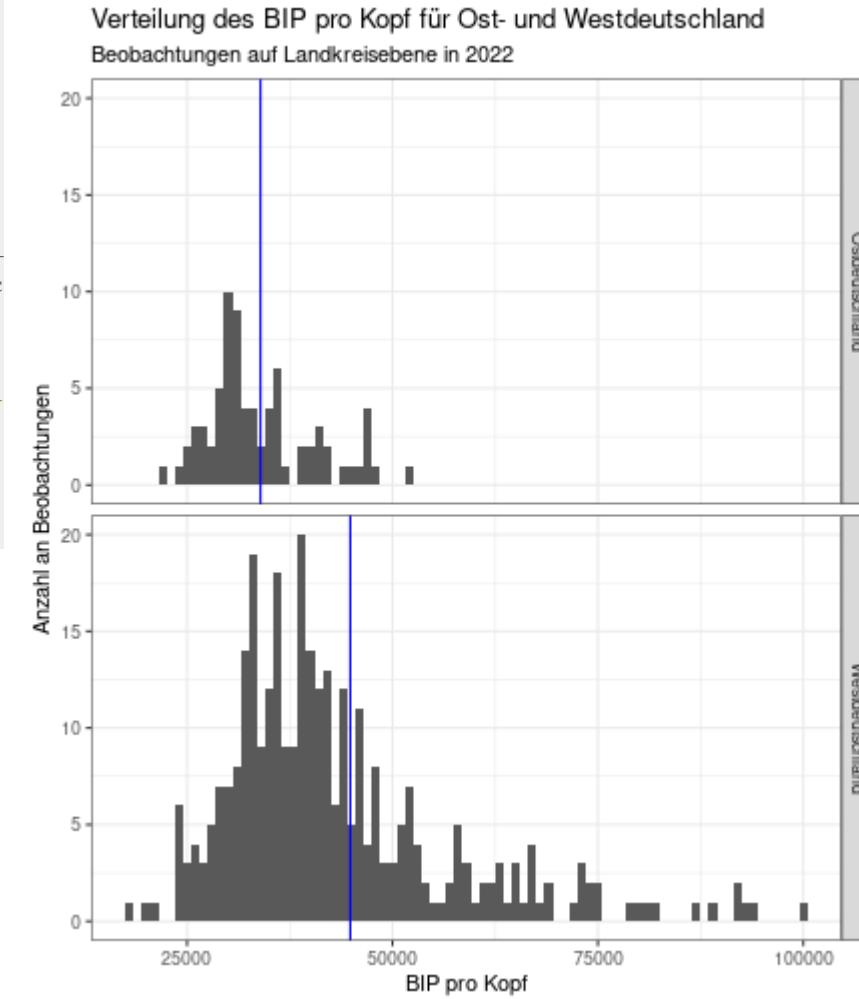


```

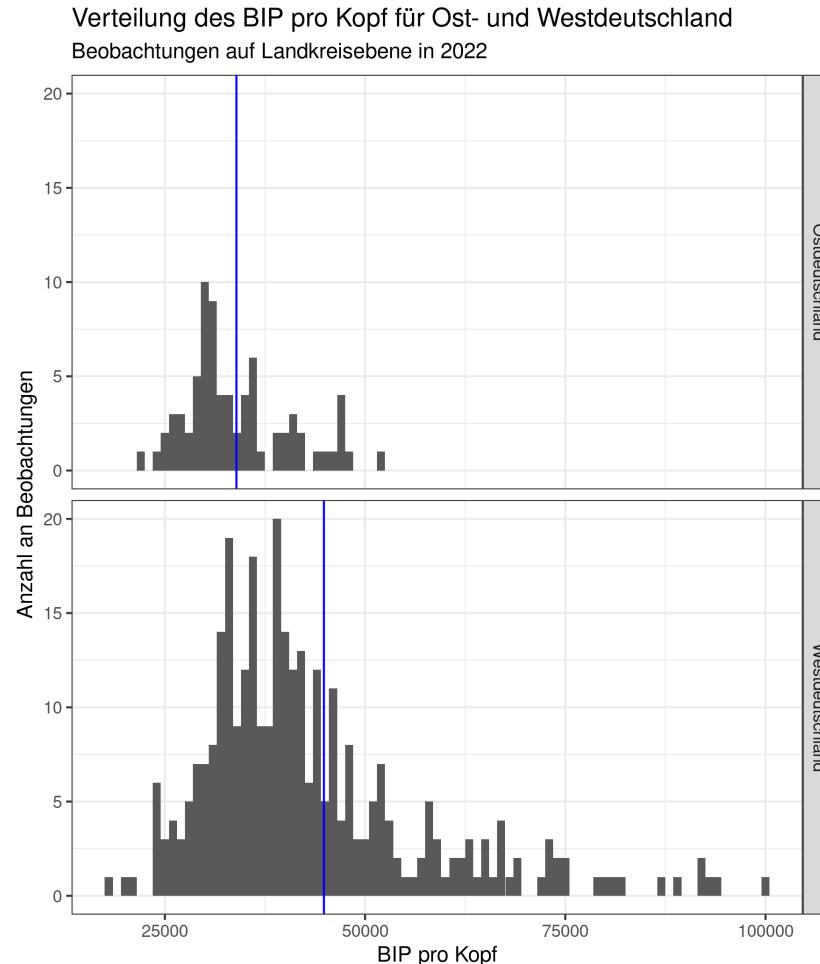
gesamtdaten %>%
  group_by(ost_name) %>%
  summarise(durchschnitt = mean(bip_pro_kopf, na.rm = TRUE))
  ungroup() -> mittel

ggplot(gesamtdaten, aes(x = bip_pro_kopf)) +
  geom_histogram(data = filter(gesamtdaten, bip_pro_kopf <= 100000), binwidth = 1000) +
  geom_vline(data = mittel, aes(xintercept = durchschnitt))
  facet_grid(ost_name ~ .) +
  theme_bw() +
  labs(title = "Verteilung des BIP pro Kopf für Ost- und Westdeutschland",
       subtitle = "Beobachtungen auf Landkreisebene",
       x = "BIP pro Kopf",
       y = "Anzahl an Beobachtungen")

```



Verteilung des BIP pro Kopf in 2022



Verteilung des BIP pro Kopf in 2022

Das Histogramm bestätigen das Bild des Boxplots:

- + Deutliche Unterschiede zwischen ost- und westdeutschend Landkreisen in 2022
- + Deutlich mehr Ausreißer nach oben bei westdeutschen Landkreisen
- + Verteilung ist für ostdeutsche Landkreise enger um den Mittelwert für das BIP pro Kopf von 33936€
- + 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

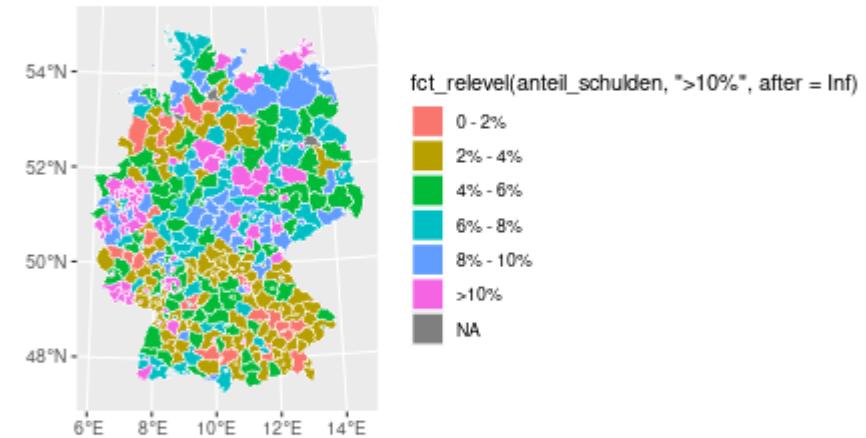
Warum könnte die Verschuldung des öffentlichen Haushalts ein Indikator für eine hohe Arbeitslosenquote sein?

Darstellung der Verschuldung der Landkreise mittels einer Deutschlandkarte.

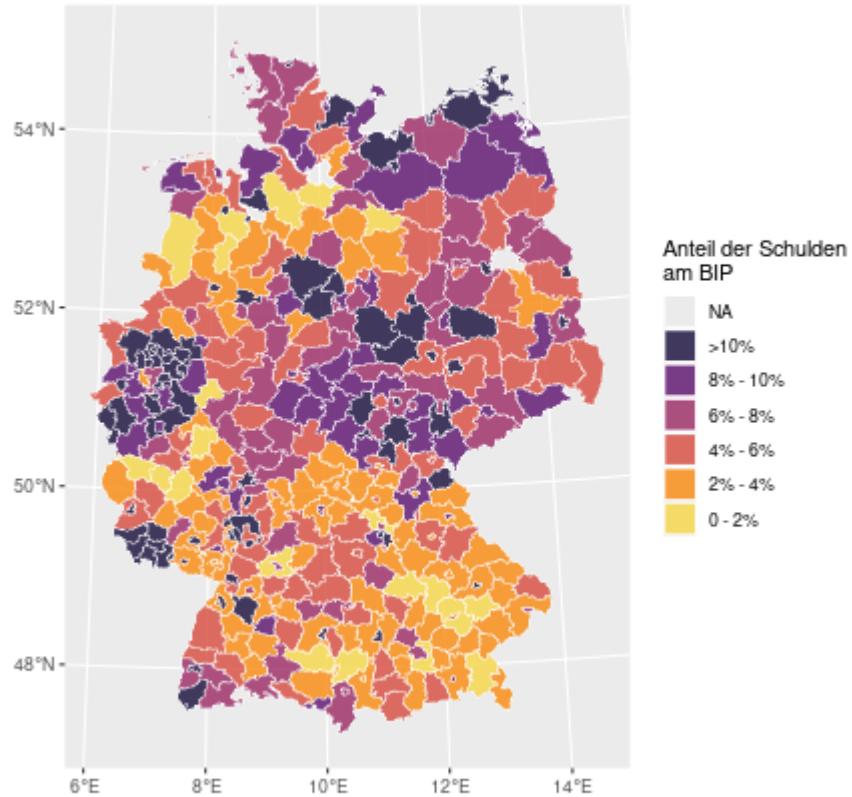
Beschreiben und interpretieren Sie die folgende Grafik.

```
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  
)
```



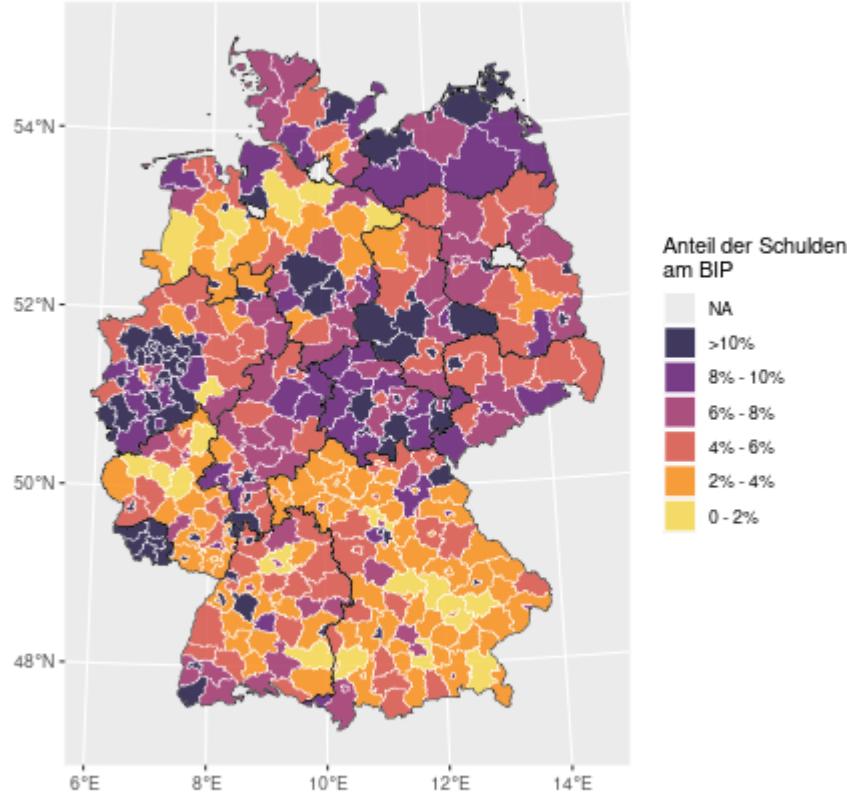
```
ggplot(  
  # Datensatz  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af),  
      color = "white",  
      size = 0.1  
    ) +  
  # Viridis Farbschema  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "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
  )

```

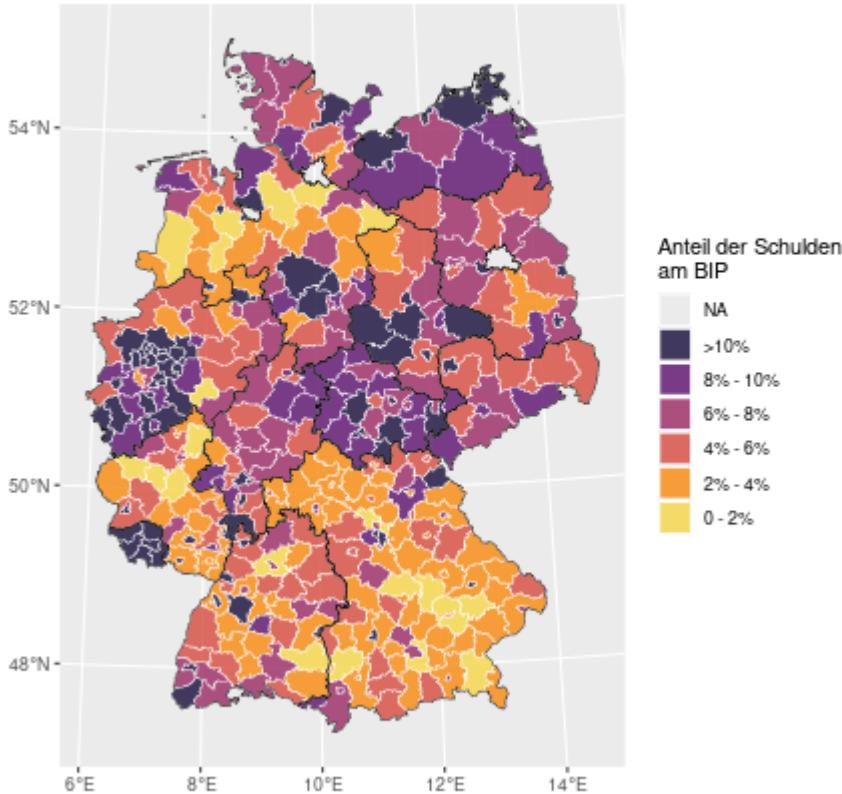


```

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 = 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 2022")

```

Wie verschuldet sind die deutschen Landkreise?
Öffentliche Schulden im Vergleich zum BIP in 2022



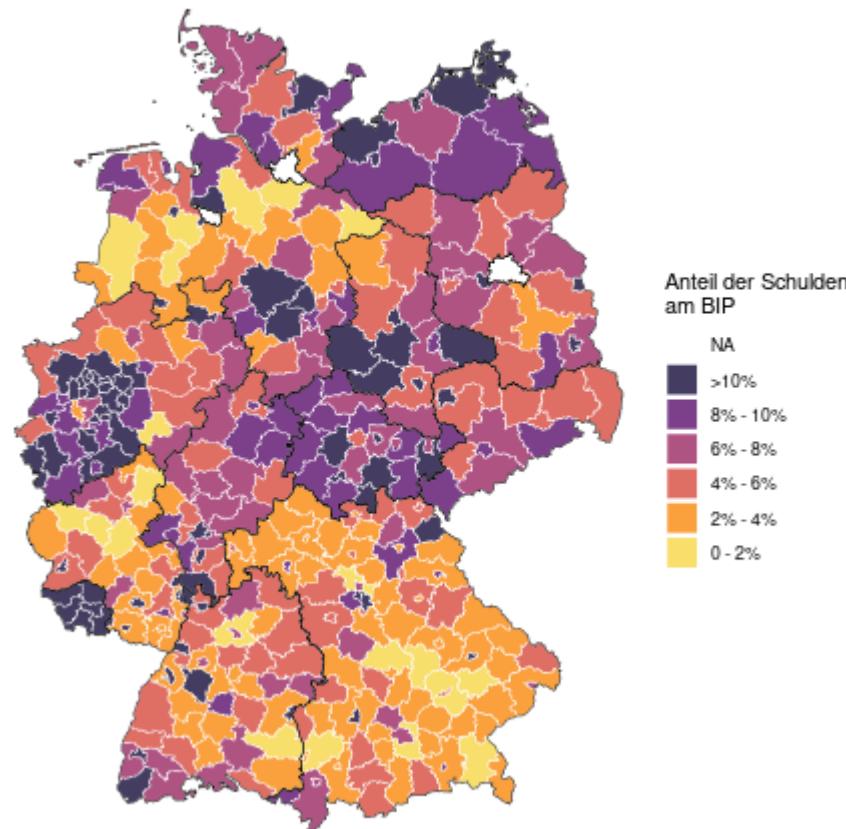
```
ggplot(  
  # Datensatz  
  data = schulden_landkreise_anteil  
) +  
  geom_sf(  
    mapping = aes(  
      fill = fct_relevel(anteil_schulden, ">10%", af  
    ),  
    color = "white",  
    size = 0.1  
) +  
  # Viridis Farbschema  
  scale_fill_viridis_d(  
    option = "inferno",  
    name = "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_void() -> plot_schulden_lk
```

```

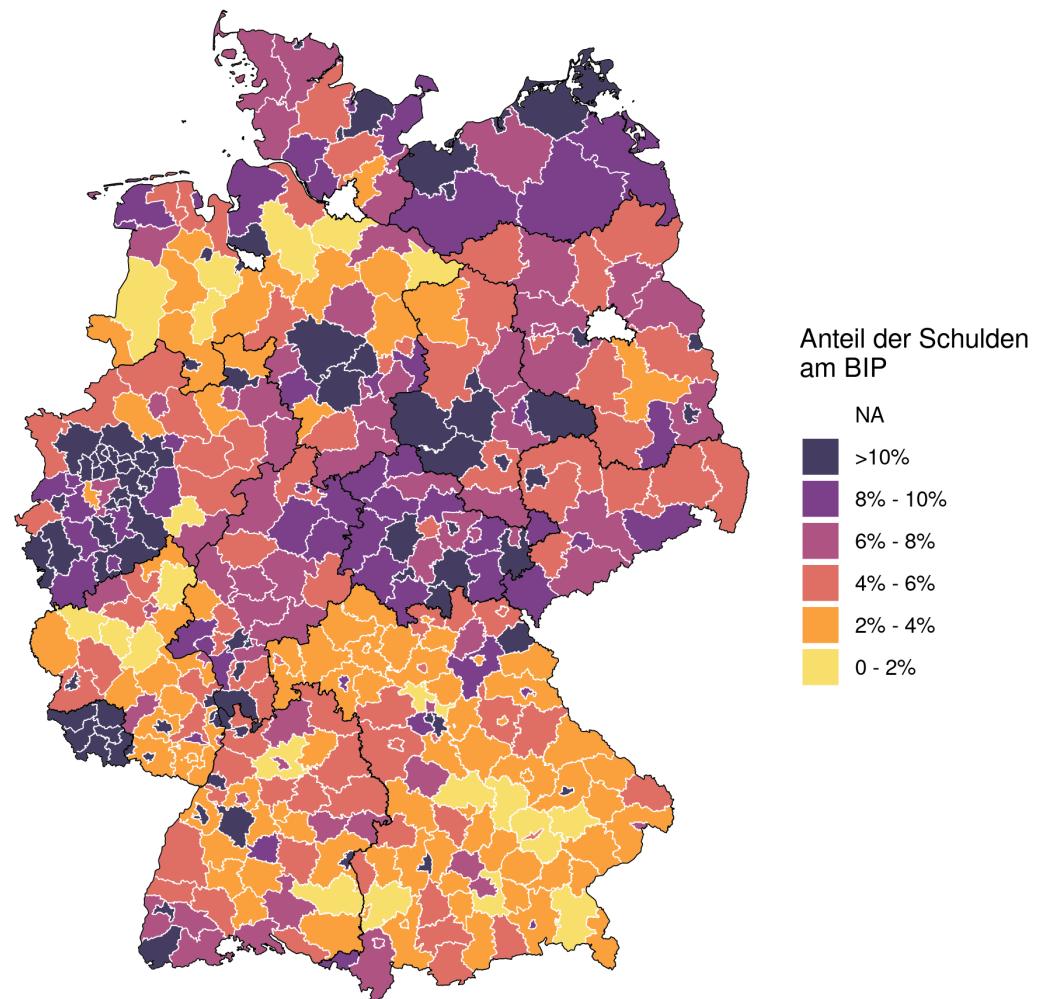
ggplot(
# Datensatz
  data = schulden_landkreise_anteil
) +
  geom_sf(
    mapping = aes(
      fill = fct_relevel(anteil_schulden, ">10%", af),
      color = "white",
      size = 0.1
    ) +
# Viridis Farbschema
  scale_fill_viridis_d(
    option = "inferno",
    name = "Anteil der Schulden\nam BIP",
    alpha = 0.8, # Deckkraft der Füllung
    begin = 0.1,
    end = 0.9,
    direction = -1,
    guide = guide_legend(reverse = T)) +
# etwas dickere Linien für Bundeslandgrenzen
  geom_sf(
    data = bundesland,
    fill = "transparent",
    color = "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 2022",
       theme_void() -> plot_schulden_lk
)

```

Wie verschuldet sind die deutschen Landkreise?
Öffentliche Schulden im Vergleich zum BIP in 2022



Wie verschuldet sind die deutschen Landkreise?
Öffentliche Schulden im Vergleich zum BIP in 2022



Verschuldung

Beschreibung:

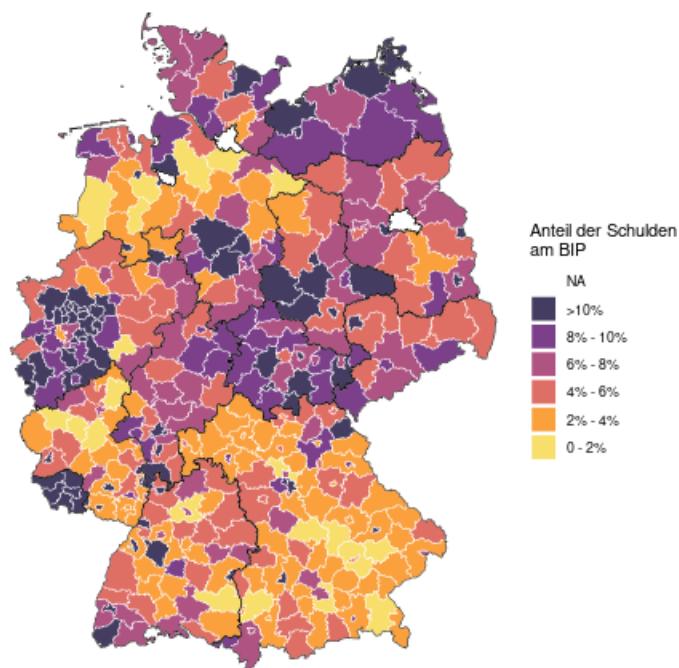
- + Niedrige Verschuldung im Verhältnis zum BIP: Baden-Württemberg, Bayern, Sachsen und Niedersachsen
- + Mittlere Verschuldung: Rheinland-Pfalz, Brandenburg, Hessen, Schleswig-Holstein
- + Hohe Verschuldung: Sachsen-Anhalt, Thüringen, Nordrhein-Westfalen, Saarland, Mecklenburg-Vorpommern

Interpretation:

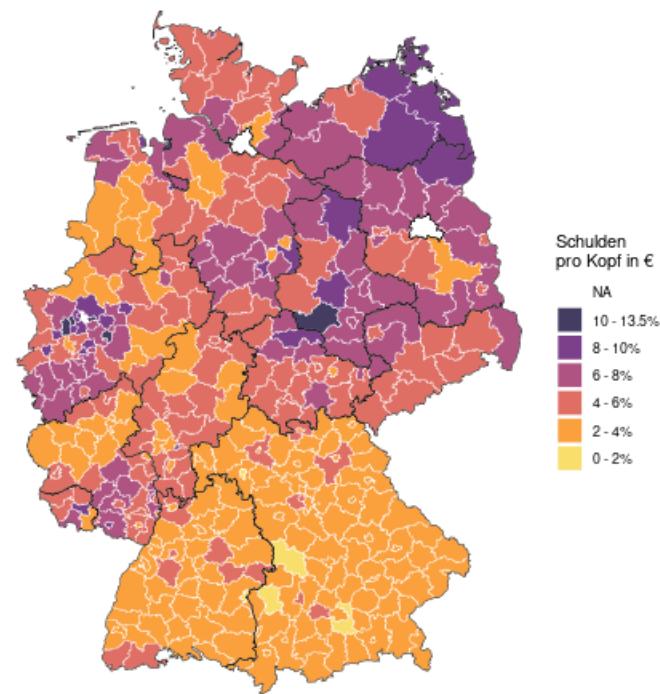
- + Strukturschwache Landkreise sind vermehrt in Ostdeutschland zu finden, allerdings scheint es eher ein Nord/Süd Gefälle als ein Ost/West Gefälle zu geben
- + Die ehemalige Herzammer der deutschen Industrie, das Ruhrgebiet, leidet unter dem Strukturwandel hin zu erneuerbaren Energien
 - + Es fallen 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 2022



Arbeitslosigkeit in Deutschland
Dargestellt ist die Arbeitslosenquote für alle Landkreise in 2022



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!

Aufgabe - Eigene Grafik zum Einkommen erstellen

Ziel: Erstellen und interpretieren Sie effektive Visualisierungen der Einkommensverteilung in Deutschland.

Ablauf:

- + Gruppenbildung (4-5 Personen): Rollen vergeben - wer kümmert sich um was? (Prompting, Coding, Grafik auswählen, Upload,...).
- + Datenauswahl & Analyse: Welche Information möchten Sie visualisieren? (Zeitreihe oder Querschnitt aus 2022?)
- + Visualisierung: Was ist die geeignete Grafik für ihre Fragestellung? Lassen Sie sich von [R Graph Gallery](#) inspirieren und erstellen Sie die für Sie passende Grafik in R (mittels ggplot2)
 - + Nutzen Sie die [R Graph Gallery](#) und [bwGPT](#) für die Erstellung
- + Interpretation der Grafik: Diskutieren Sie die Grafik in der Gruppe
- + Upload: Code für Grafik + kurze Interpretation im Google Forms hochladen: <https://forms.gle/R4qJEegZnFoTHU7f7>
- + Feedback-Runde im Plenum: Anonymes Feedback von Kommiliton*innen und mir.

Materialien: Laptops mit R, RStudio, bereitgestellter Datensatz zu Einkommensdaten

Erkenntnisse: Effektive Grafiktypen auswählen und interpretieren + Teamarbeit stärken.

30 : 00

Bivariate deskriptive Analyse

Die Korrelation

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

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

Hierzu nutzen wir die Korrelation der Variablen!

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

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

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

Die Korrelation

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

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

Hierzu nutzen wir die Korrelation der Variablen!

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

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

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

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

Korrelation zwischen Arbeitslosenquote und BIP-pro-Kopf-Wachstum

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

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

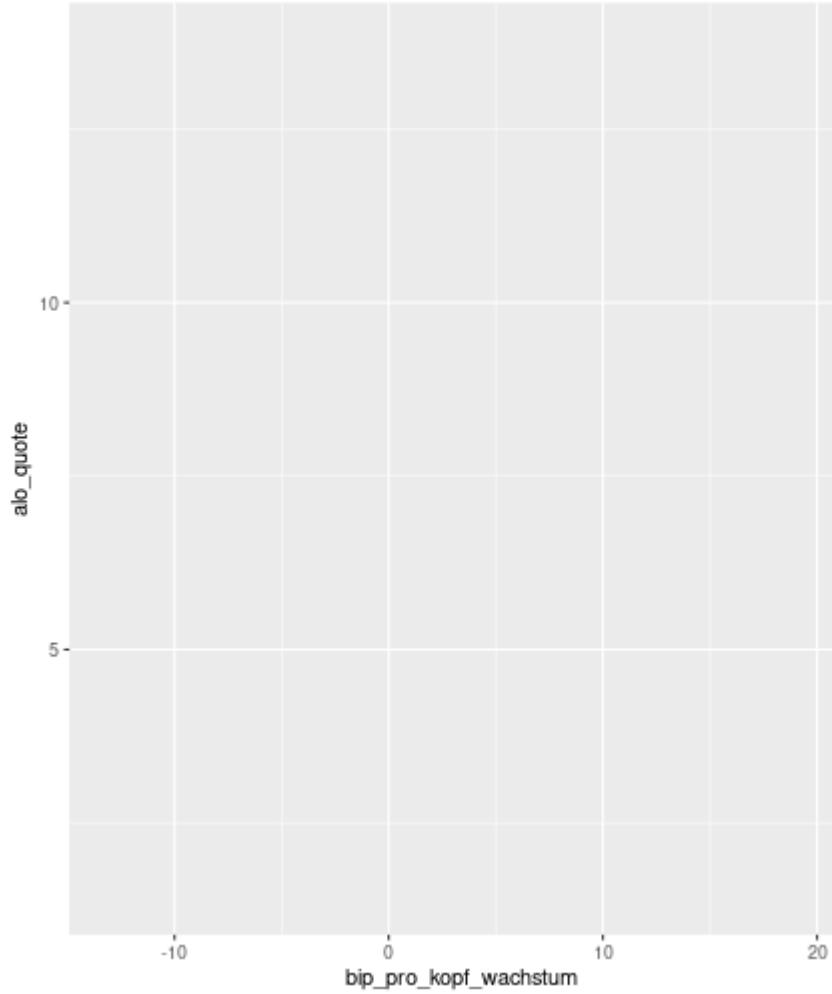
Im Streudiagramm können wir Variable x auf der x-Achse und Variable y auf der y-Achse abtragen

gesamtdaten

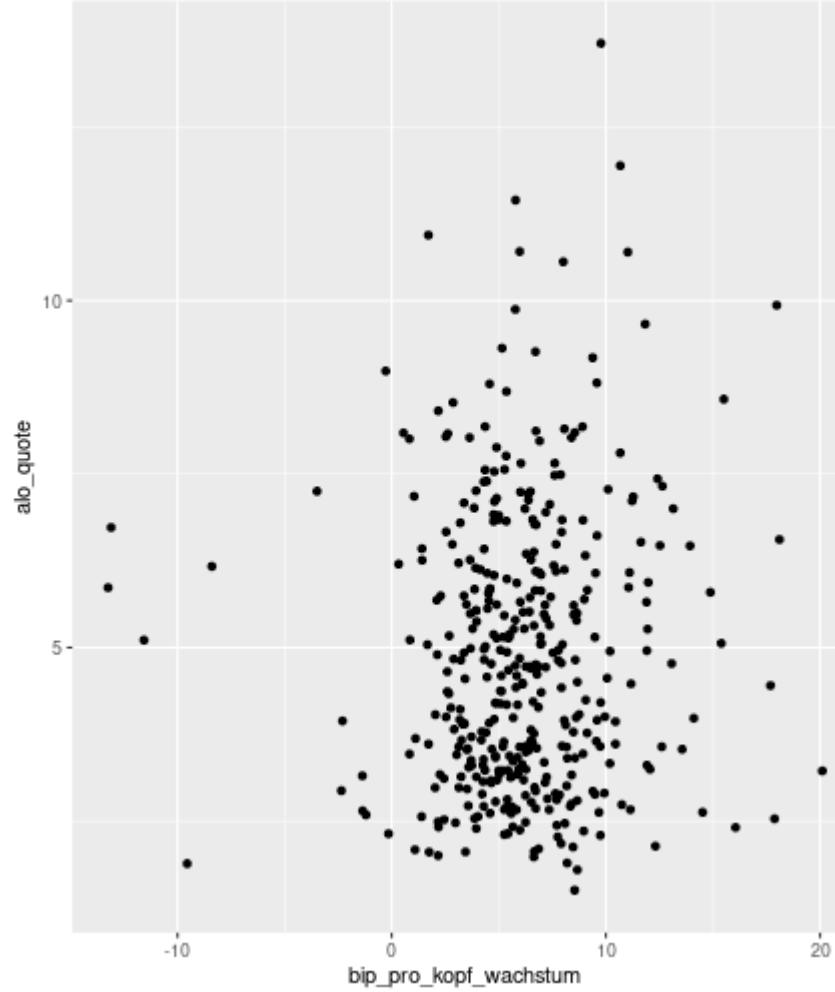
```
## # A tibble: 400 × 17
##   Regionalschluessel total_alo landkreis_name      bundesland bundesland_name
##   <chr>              <dbl> <chr>                <chr>      <chr>
## 1 01001                  3970. Flensburg          01        Schleswig-Holst
## 2 01002                  10315. Kiel             01        Schleswig-Holst
## 3 01003                  8776. Lübeck           01        Schleswig-Holst
## 4 01004                  3359. Neumünster       01        Schleswig-Holst
## 5 01051                  3858. Dithmarschen      01        Schleswig-Holst
## 6 01053                  5351. Herzogtum Lauenburg 01        Schleswig-Holst
## 7 01054                  4155. Nordfriesland     01        Schleswig-Holst
## 8 01055                  4824. Ostholstein      01        Schleswig-Holst
## 9 01056                  8547. Pinneberg         01        Schleswig-Holst
## 10 01057                 2572. 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>
```

```
gesamtdaten %>%
```

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

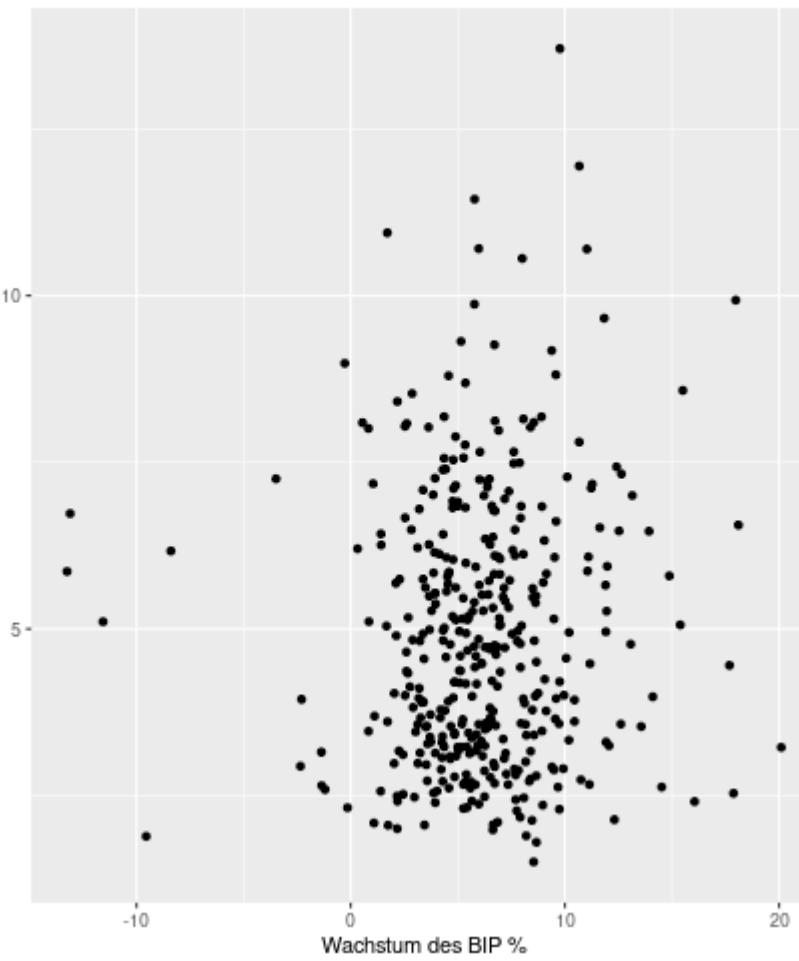


```
gesamtdaten %>%  
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot  
  geom_point()
```



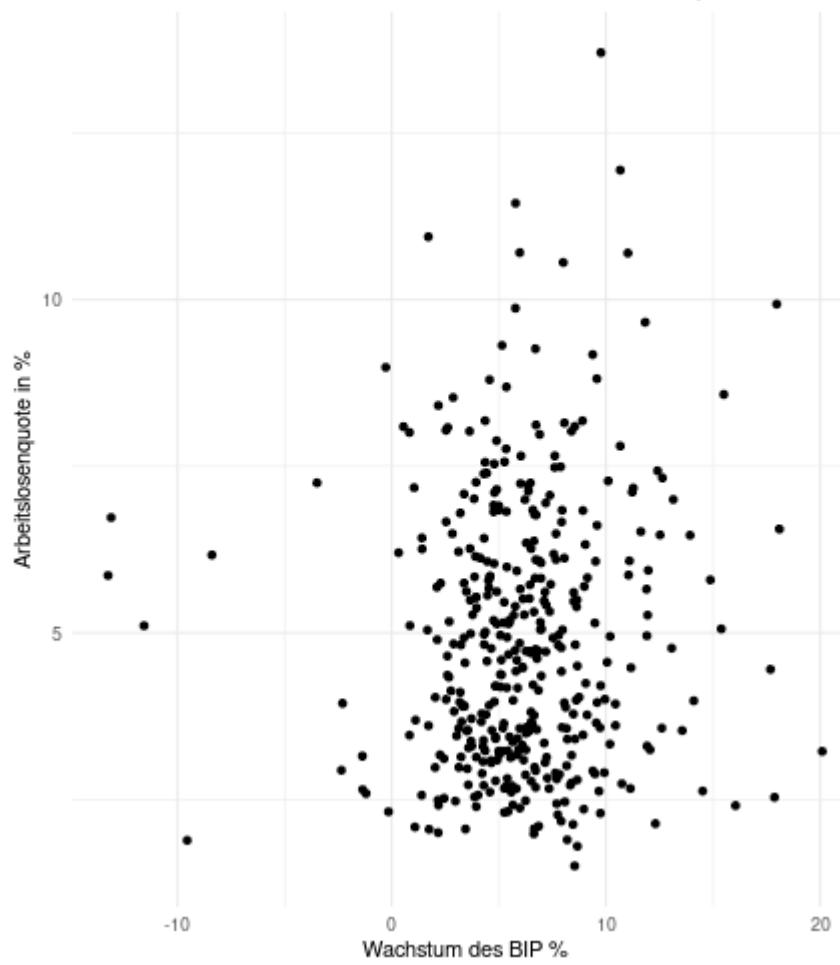
```
gesamtdaten %>%  
  ggplot(aes(x = bip_pro_kopf_wachstum, y = alo_quot  
  geom_point() +  
  labs( x = "Wachstum des BIP %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation des BIP-Wachstums und d
```

Korrelation des BIP-Wachstums und der Arbeitslosenquote



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

Korrelation des BIP-Wachstums und der Arbeitslosenquote



Korrelation zwischen Arbeitslosenquote und BIP-Wachstum

- + Es fallen die Ausreißer ins Auge (+20% und -12%)
 - + Vorheriges Jahr hohes/niedriges BIP, dadurch jetzt niedriges/hohes BIP-Wachstum
 - + Die Energiekrise war bei den negativen Wachstumsraten der Auslöser (Ludwigshafen mit BASF war der betroffene Landkreis)
- + 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] 0.05423851
```

Korrelation zwischen Arbeitslosenquote und BIP-Wachstum

- + Es fallen die Ausreißer ins Auge (+20% und -12%)
 - + Vorheriges Jahr hohes/niedriges BIP, dadurch jetzt niedriges/hohes BIP-Wachstum
 - + Die Energiekrise war bei den negativen Wachstumsraten der Auslöser (Ludwigshafen mit BASF war der betroffene Landkreis)
- + 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")
```

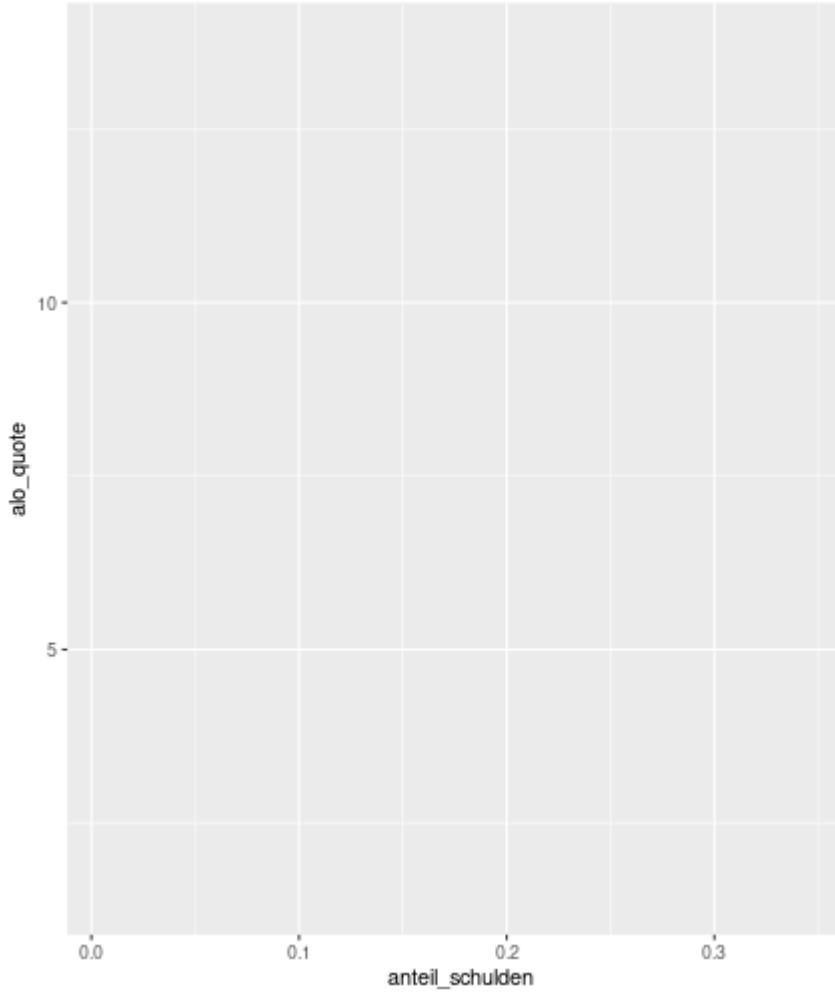
```
## [1] 0.05423851
```

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

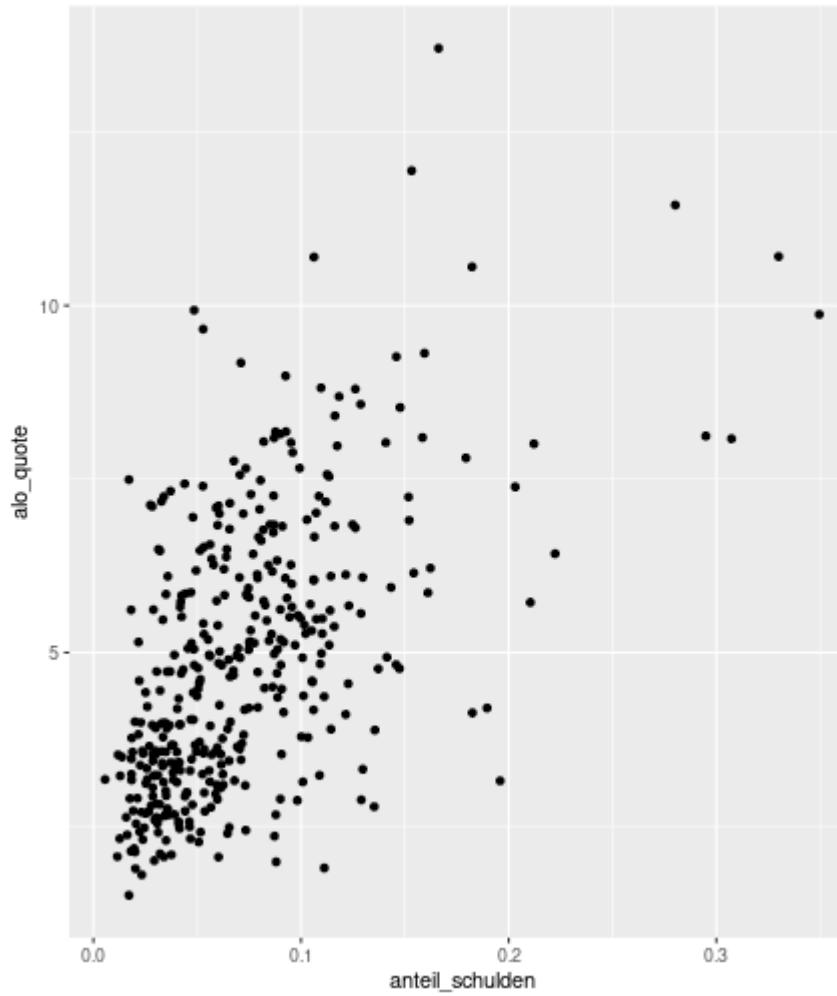
```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g
```

```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten  
  
## # A tibble: 400 × 17  
##   Regionalschluessel total_alo landkreis_name      bundesland bundesland_name  
##   <chr>           <dbl> <chr>          <chr>        <chr>  
## 1 01001            3970. Flensburg       01 Schleswig-Holst  
## 2 01002            10315. Kiel          01 Schleswig-Holst  
## 3 01003            8776. Lübeck         01 Schleswig-Holst  
## 4 01004            3359. Neumünster     01 Schleswig-Holst  
## 5 01051            3858. Dithmarschen    01 Schleswig-Holst  
## 6 01053            5351. Herzogtum Lauenburg 01 Schleswig-Holst  
## 7 01054            4155. Nordfriesland    01 Schleswig-Holst  
## 8 01055            4824. Ostholstein     01 Schleswig-Holst  
## 9 01056            8547. Pinneberg       01 Schleswig-Holst  
## 10 01057           2572. 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))
```

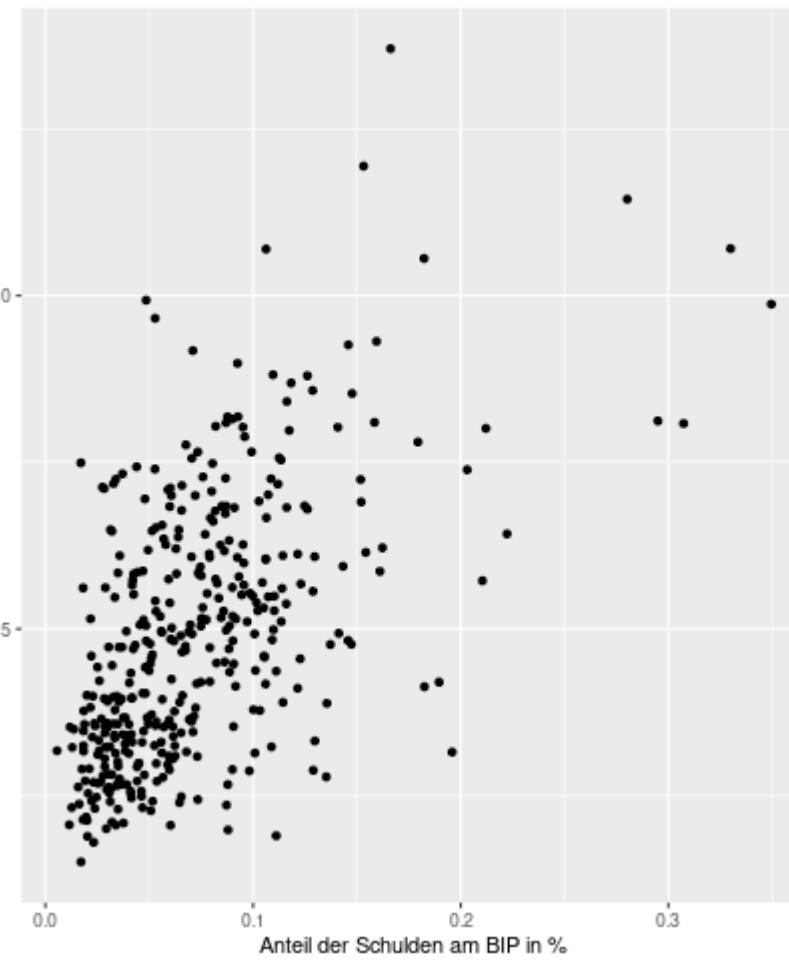


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point()
```

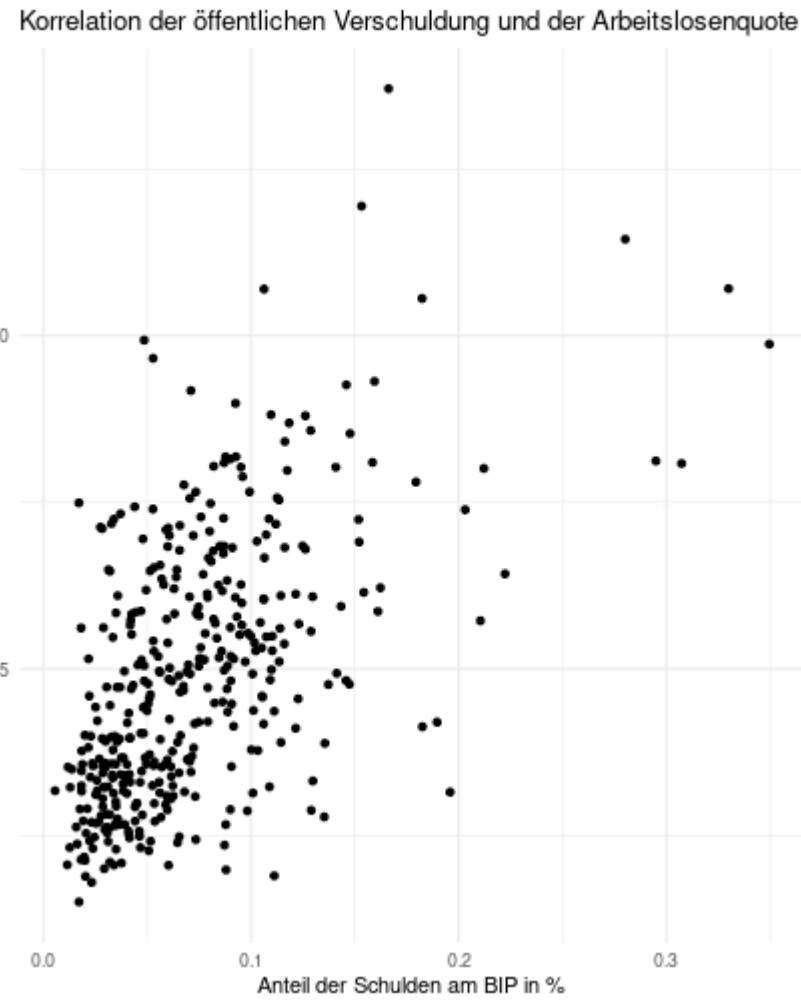


```
cor_alo_verschuldung <- cor(gesamtdaten$alo_quote, g  
gesamtdaten %>%  
  ggplot(aes(x = anteil_schulden, y = alo_quote)) +  
  geom_point() +  
  labs( x = "Anteil der Schulden am BIP in %",  
        y = "Arbeitslosenquote in %",  
        title = "Korrelation der öffentlichen Versch
```

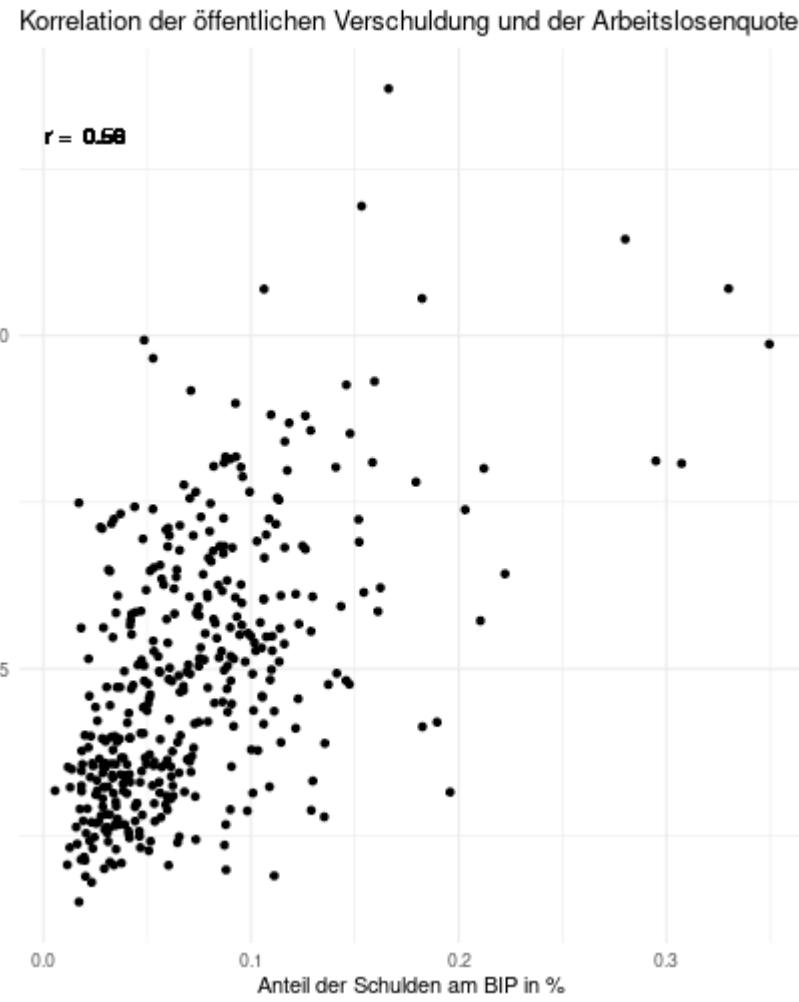
Korrelation der öffentlichen Verschuldung und der Arbeitslosenquote



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



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



Korrelation zwischen Arbeitslosenquote und Verschuldung

Hier ist der positive Zusammenhang zwischen Verschuldung (x-Achse) und Arbeitslosenquote (y-Achse) deutlicher
Korrelationskoeffizient zeigt mit $\rho = 0.56$ auch einen starken Zusammenhang

ρ Beschreibung (nährungsweise)

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

Korrelation zwischen Arbeitslosenquote und Verschuldung

Hier ist der positive Zusammenhang zwischen Verschuldung (x-Achse) und Arbeitslosenquote (y-Achse) deutlicher
Korrelationskoeffizient zeigt mit $\rho = 0.56$ auch einen starken Zusammenhang

ρ Beschreibung (nährungsweise)

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

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

Interpretation der Korrelation

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

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

Zusammenfassung und Ausblick

Dieses Semester: Deskriptiven Statistik

Nächstes Semester: Induktive Statistik, insbesondere durch lineare Regressionen

Was haben wir bisher gelernt?

Zusammenfassung und Ausblick

Dieses Semester: Deskriptiven Statistik

Nächstes Semester: Induktive Statistik, insbesondere durch lineare Regressionen

Was haben wir bisher gelernt?

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

Übungsaufgaben

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

- ✚ Erstellen Sie eine deskriptive Tabelle, welche das Einkommen für das Jahr 2022 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.