

# DataVis Examples - R

November 8, 2020

## 1 ZEWK - Hands On Datenvisualisierung, Beispiele in R

Seminar von Letty und Karen

Dieses Notebook dient als Beispiel für die Implementation einer Datenvisualisierungspipeline in der Sprache R. Es kann sowohl in Jupyter als auch in Jupyter Lab ausgeführt werden, jedoch können sich einzelne Shortcuts unterscheiden.

### 1.1 Benutzung von Jupyter (Lab)

Hier ein paar praktische und wichtige Kommandos und Tastenkombinationen die ihr kennen solltet: Außerhalb einer Zelle:

- ENTER - Zelle editieren
- strg + ENTER - Zelle ausführen
- shift + ENTER - Zelle ausführen und zur nächsten gehen

Innerhalb einer Zelle (Editiermodus der Zelle):

- ESC - Zelle verlassen
- D, D - Zelle löschen
- A - leere Zelle oberhalb (above) einfügen
- B - leere Zelle unterhalb (below) einfügen

#### 1.1.1 Setup von hilfreichen Python Packages

```
[1]: library(ggplot2)    ## graphics plotting package
```

#### 1.1.2 Import der Daten

```
[2]: my_data <- read.csv("owid-covid-data.csv", header=TRUE, sep = ",")
my_data['date'] <- as.Date(my_data$date) # Umwandeln von string format zu
↳ datetime format

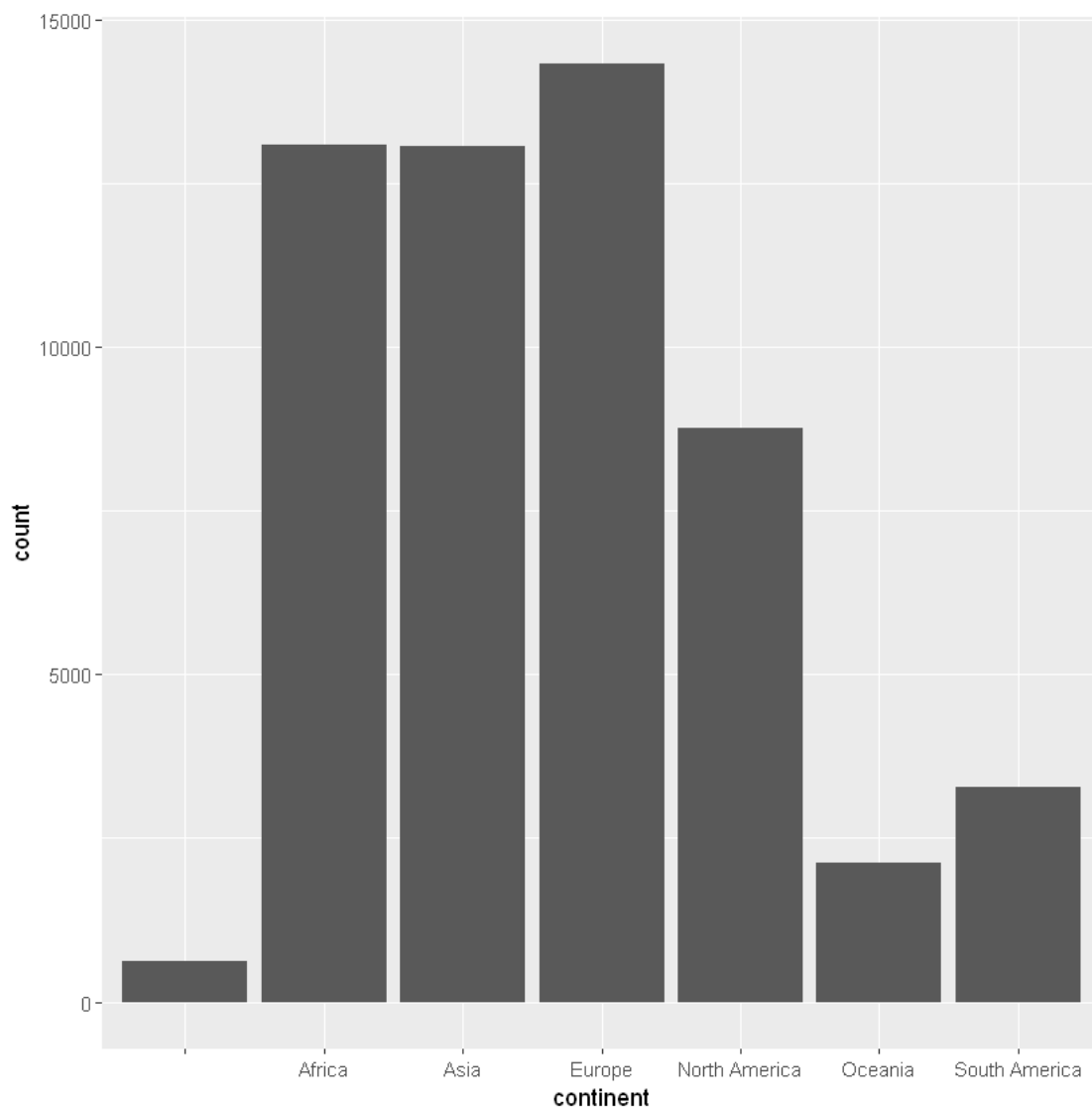
# Reduktion der Datenmenge - beispiel daten für GDP aller Länder in Europa
```

```
data_sample <- my_data[!duplicated(my_data$location), ]
data_gdp = data_sample[,c('iso_code', 'continent', 'location', '
  → 'population_density', 'median_age', 'aged_65_older',
    'aged_70_older', 'gdp_per_capita', 'extreme_poverty', 'life_expectancy')]
data_gdp <- data_gdp[data_gdp$continent == "Europe", ]

data_austria <- my_data[my_data$location == "Austria", ]
```

## 2 Bar chart

```
[3]: # wie viele einträge haben wir pro kontinent? als Graphik
ggplot(data = my_data) + geom_bar(mapping = aes(x = continent))
```



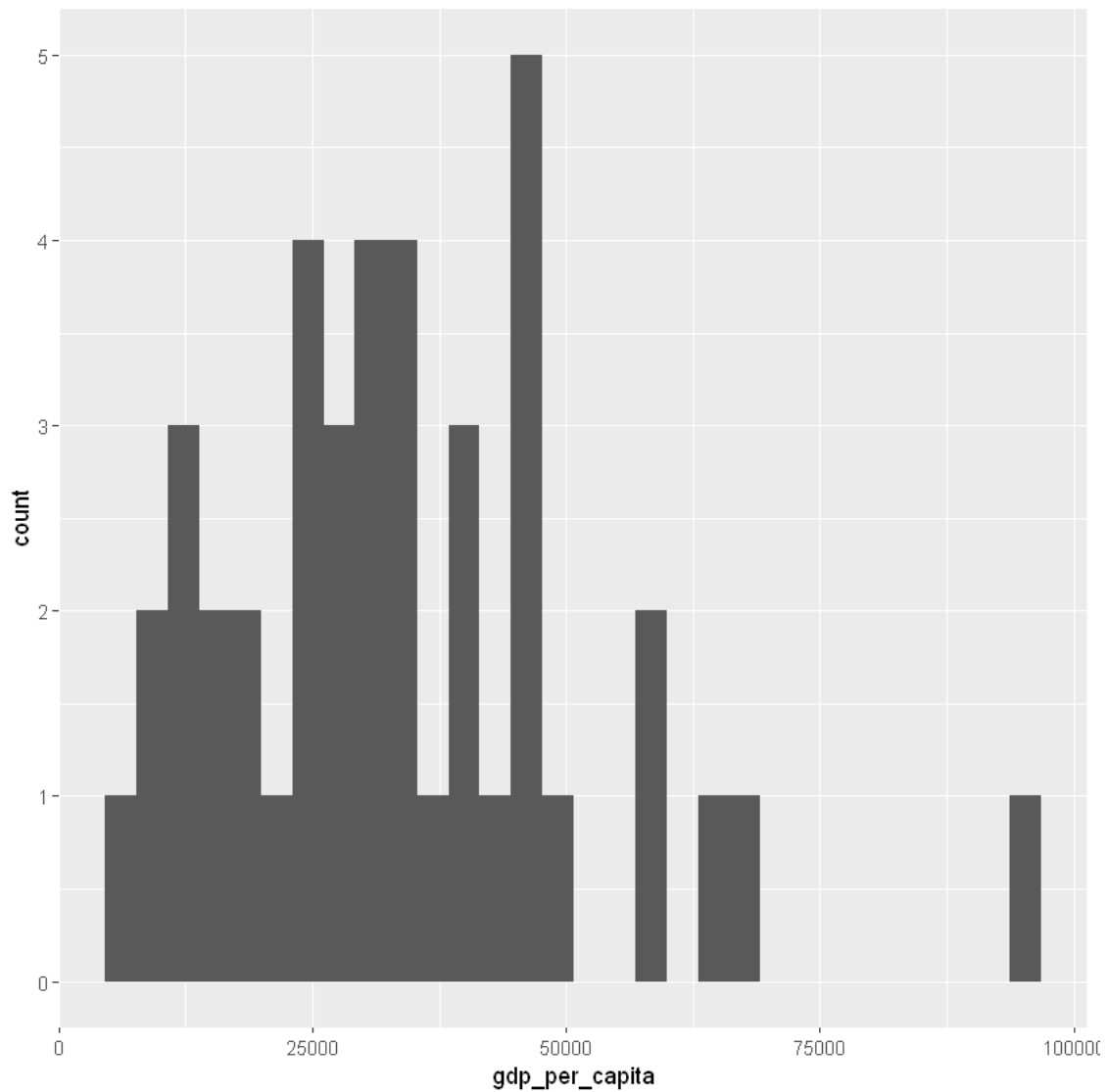
### 3 Histogram

```
[4]: ggplot(data_gdp, aes(x=gdp_per_capita)) + geom_histogram()
```

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning message:

"Removed 9 rows containing non-finite values (stat\_bin)."

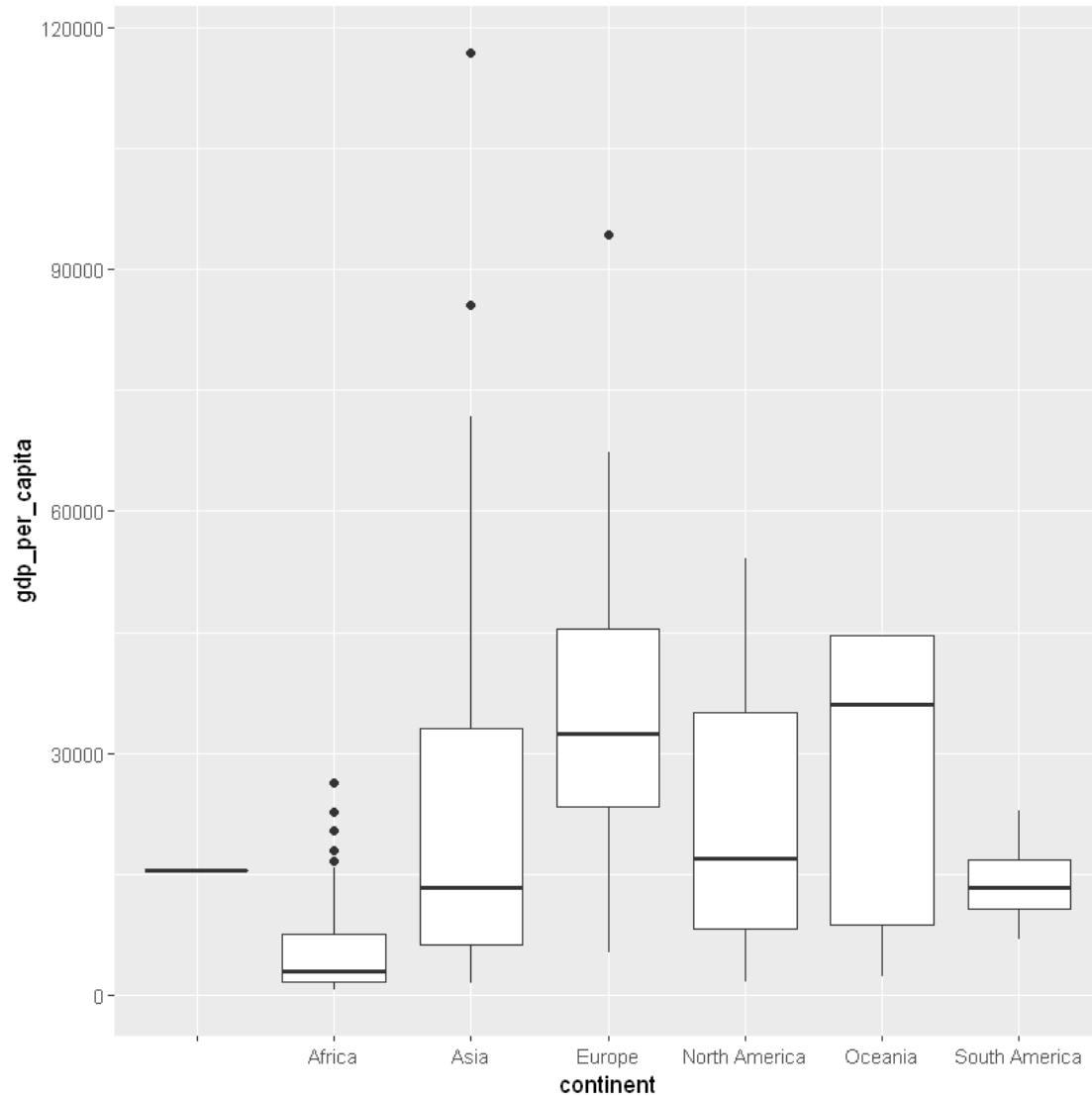


## 4 Boxplot

```
[5]: ggplot(my_data, aes(x=continent, y=gdp_per_capita)) + geom_boxplot(mapping =  
  ↪ aes(x = continent))
```

Warning message:

"Removed 6737 rows containing non-finite values (stat\_boxplot)."

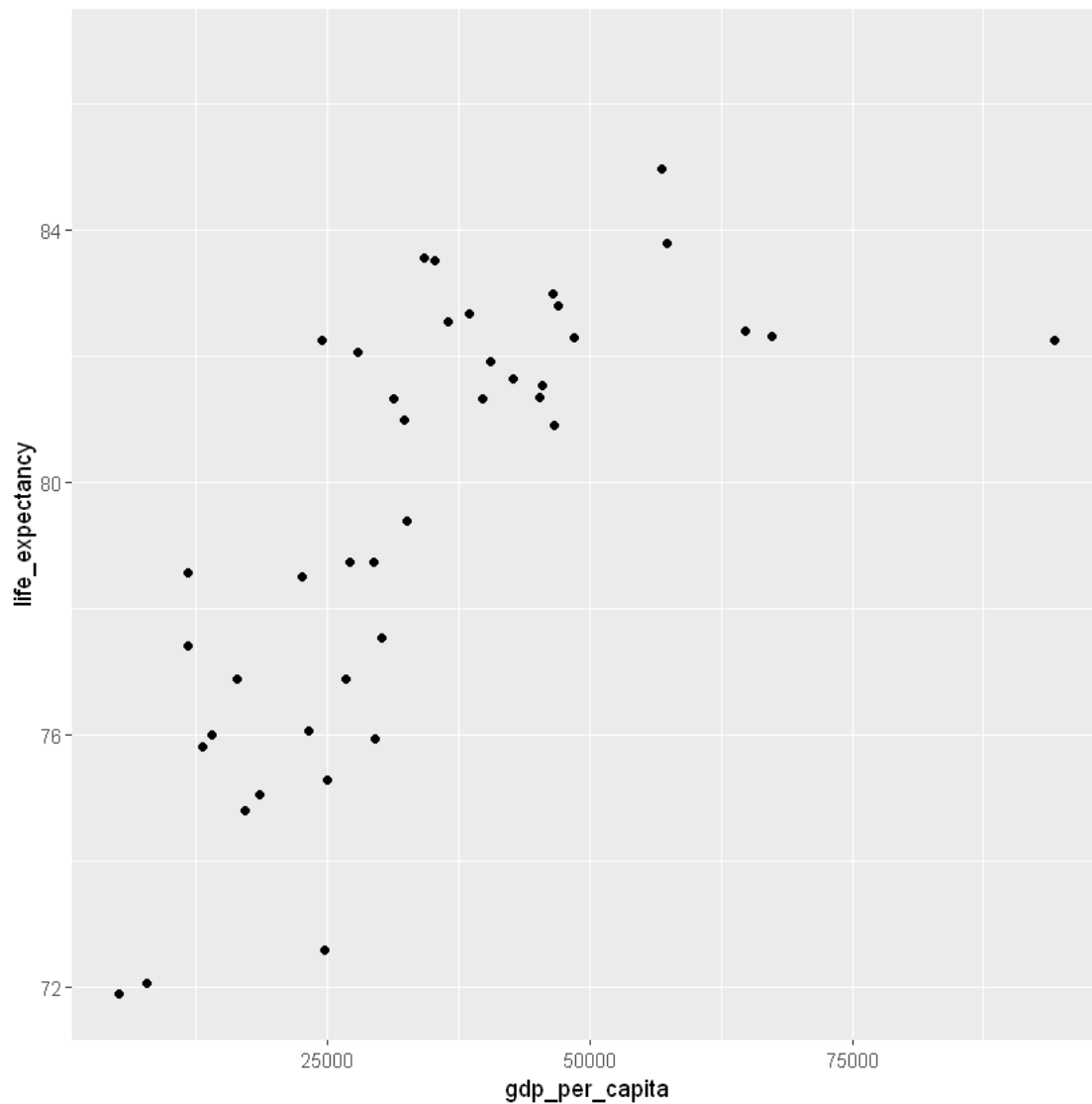


## 5 Scatterplot

```
[6]: ggplot(data_gdp, aes(x=gdp_per_capita, y=life_expectancy)) + geom_point()
```

Warning message:

"Removed 10 rows containing missing values (geom\_point)."

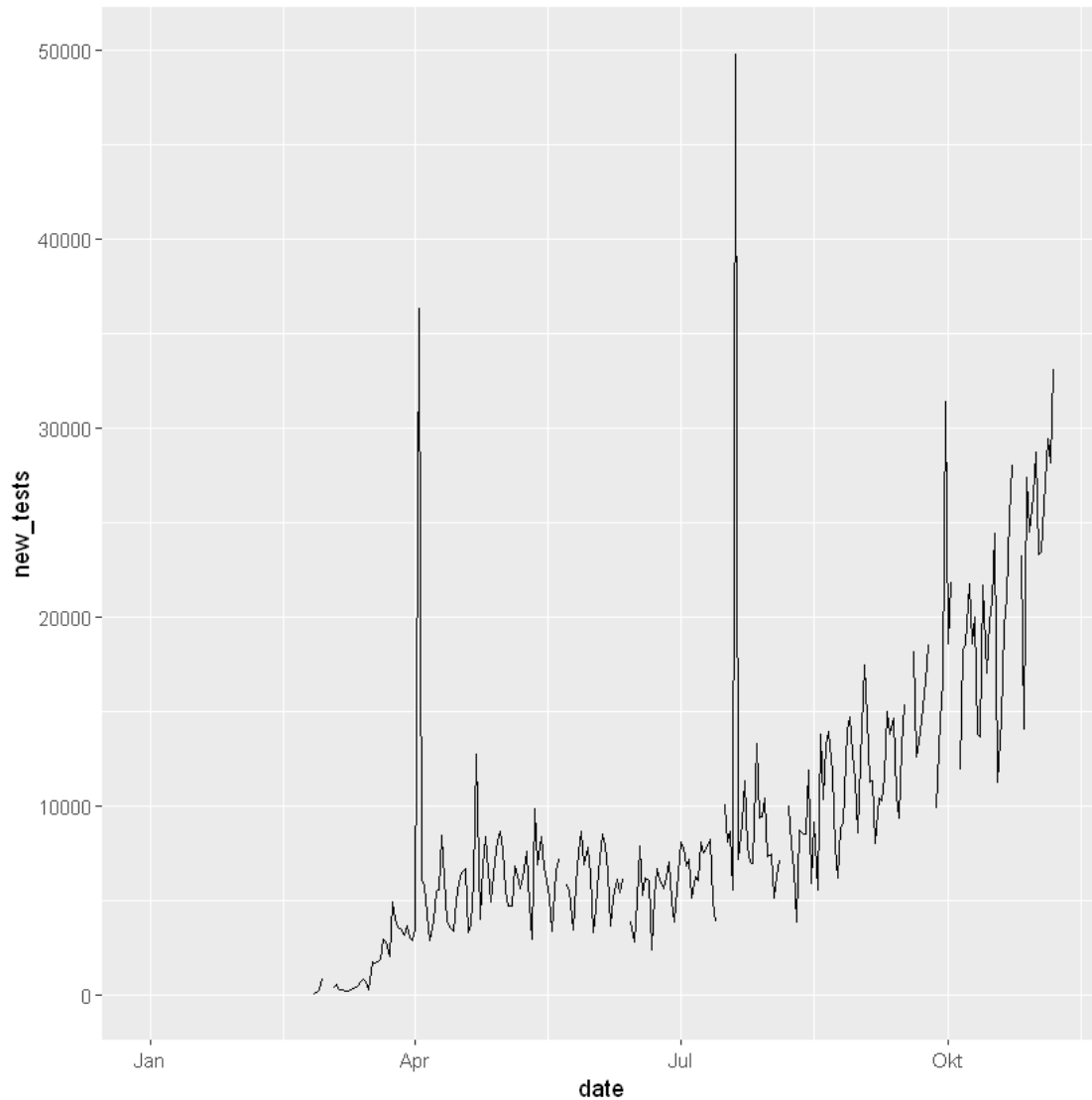


## 6 Line chart

```
[7]: ggplot(data_austria, aes(x=date, y=new_tests)) + geom_line()
```

Warning message:

"Removed 58 row(s) containing missing values (geom\_path)."

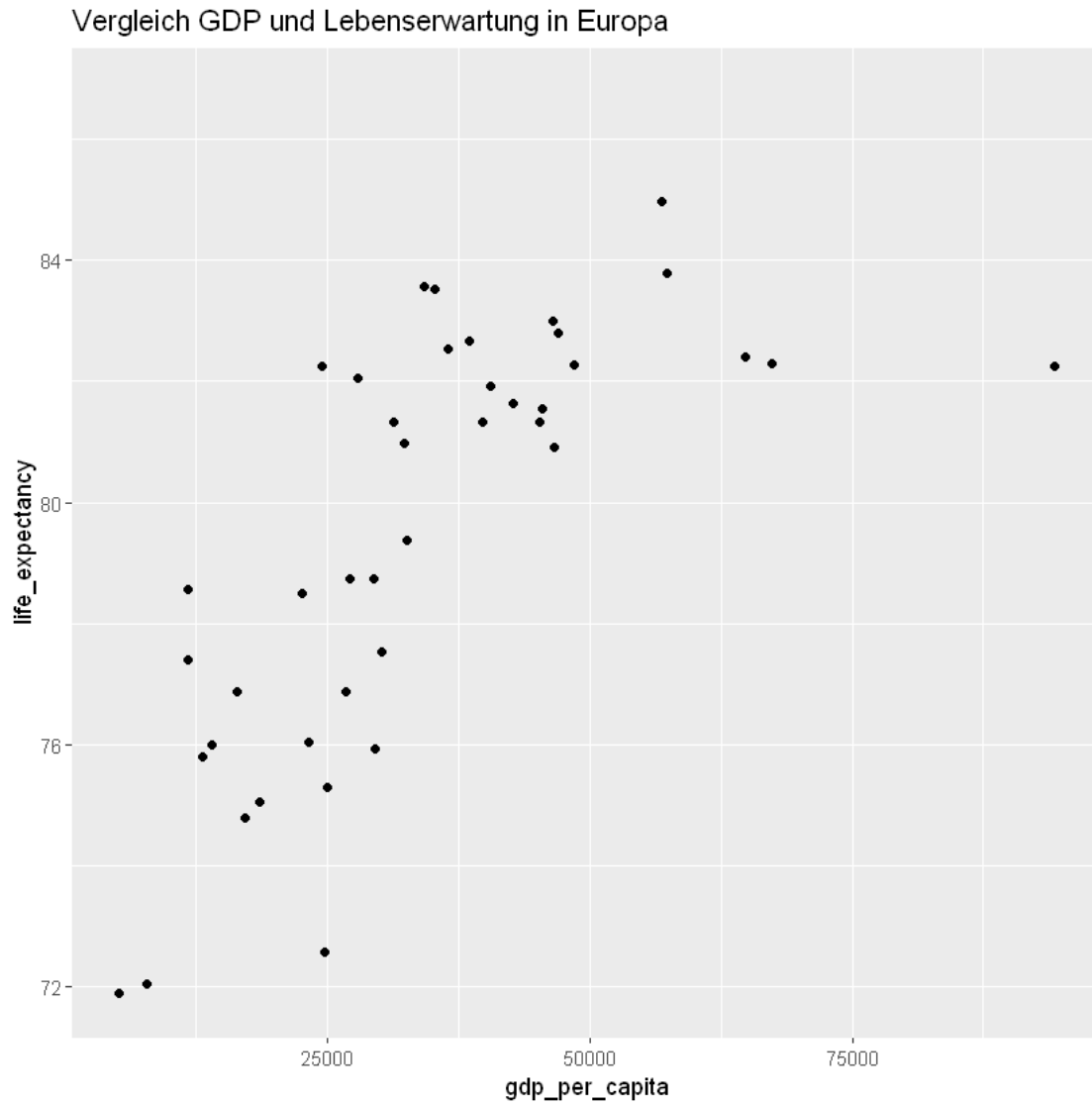


## 7 Titel und Achsenbeschriftung

```
[8]: ggplot(data_gdp, aes(x=gdp_per_capita, y=life_expectancy)) + geom_point() +  
      ggtitle("Vergleich GDP und Lebenserwartung in Europa")
```

Warning message:

"Removed 10 rows containing missing values (geom\_point)."

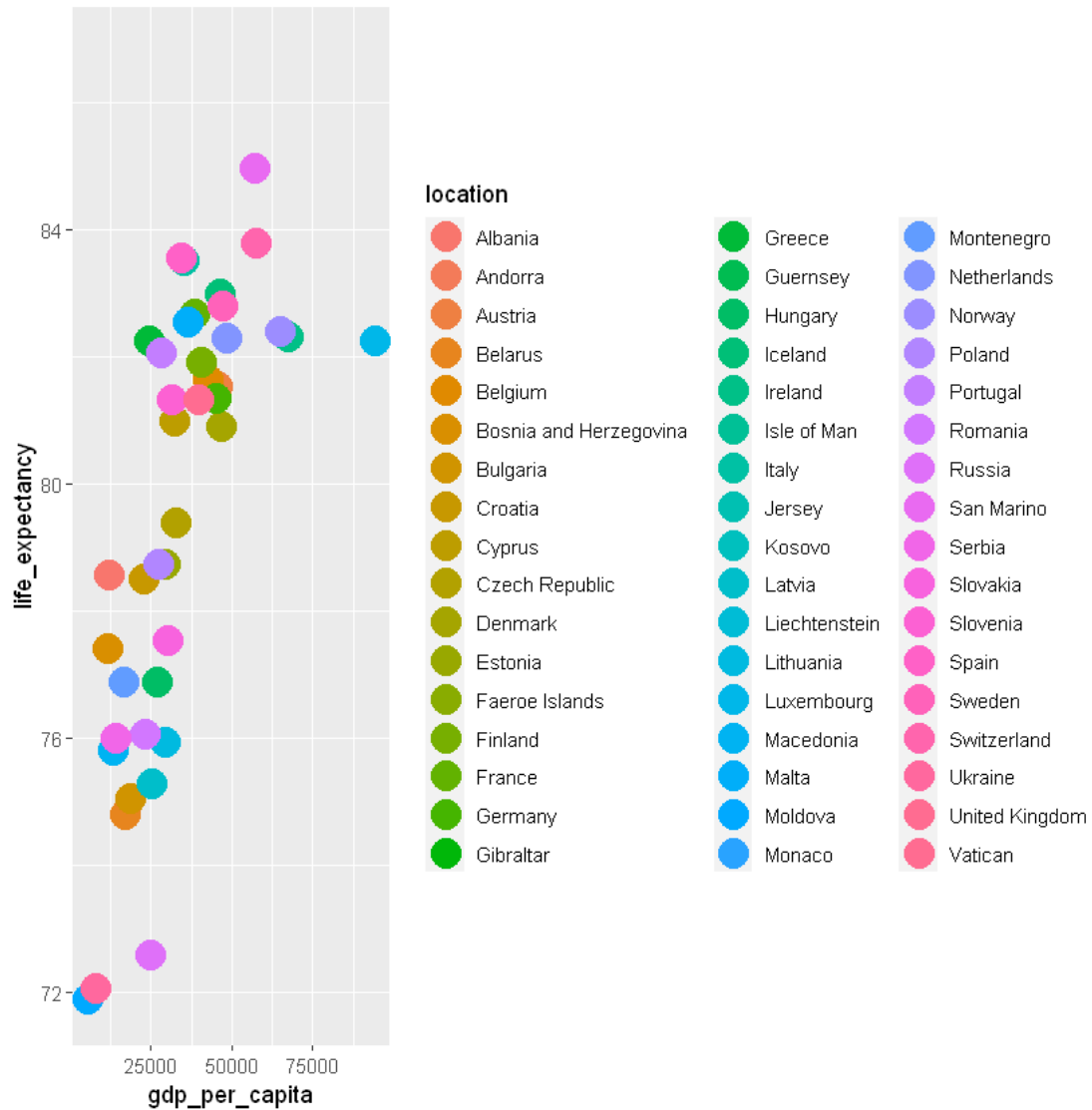


## 8 Farbänderungen

```
[9]: ggplot(data_gdp, aes(x=gdp_per_capita, y=life_expectancy, color=location)) +  
      ↪ geom_point(size=6)
```

Warning message:

"Removed 10 rows containing missing values (geom\_point)."

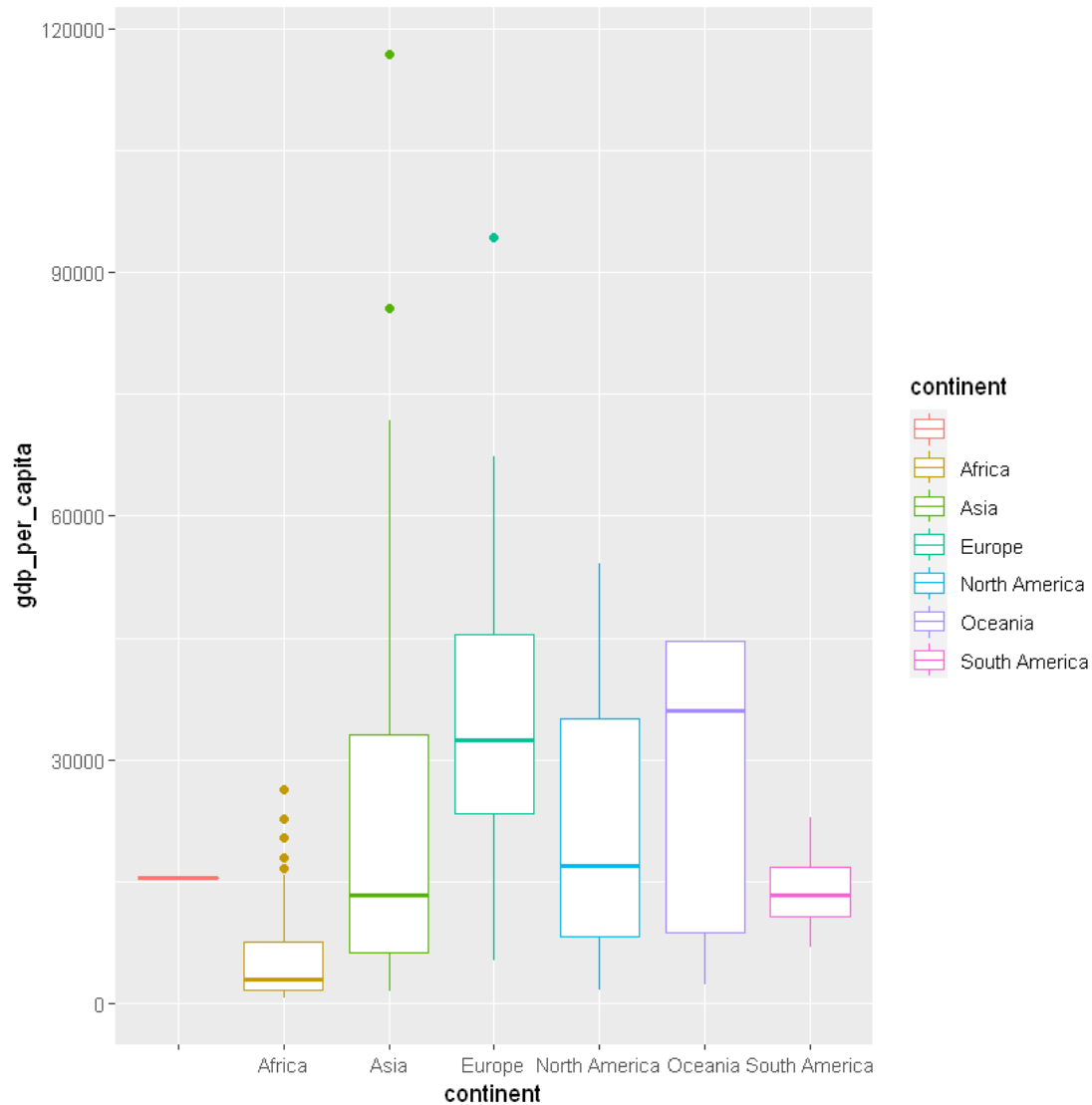


```
[10]: ggplot(my_data, aes(x=continent, y=gdp_per_capita, color=continent)) +  
      geom_boxplot(mapping = aes(x = continent))
```

Warning message:

"Removed 6737 rows containing non-finite values (stat\_boxplot)."





## 9 Export

```
[11]: ggplot(data_gdp, aes(x=gdp_per_capita, y=life_expectancy)) + geom_point() +  
      ggtitle("Vergleich GDP und Lebenserwartung in Europa")  
      ggsave("R - scatterplot example.svg")
```

Warning message:

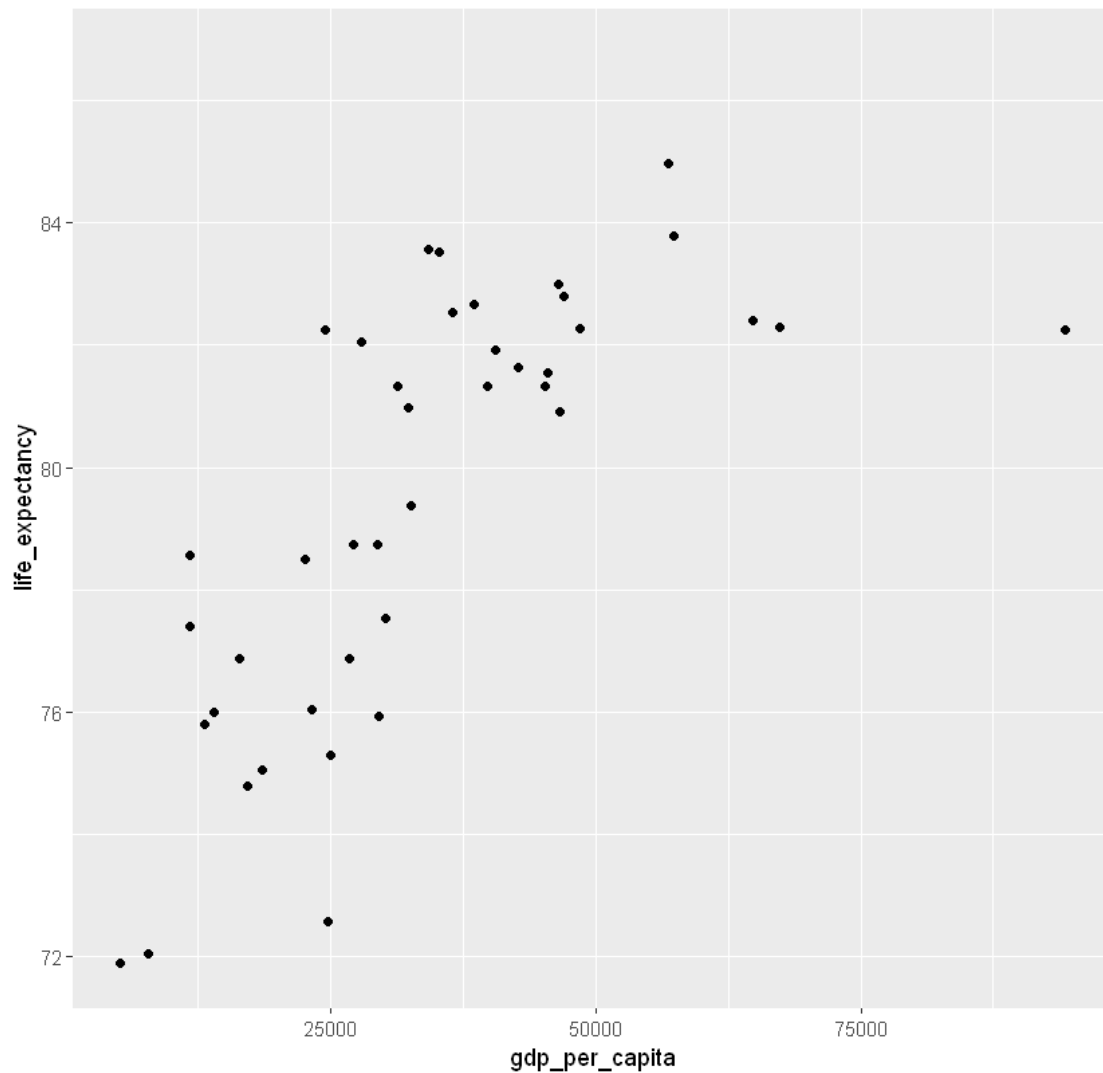
"Removed 10 rows containing missing values (geom\_point)."

Saving 6.67 x 6.67 in image

Warning message:

"Removed 10 rows containing missing values (geom\_point)."

### Vergleich GDP und Lebenserwartung in Europa



```
[12]: ggplot(my_data, aes(x=continent, y=gdp_per_capita, color=continent)) +  
  ↪geom_boxplot(mapping = aes(x = continent))  
  ggsave("R - Boxplot example.svg")
```

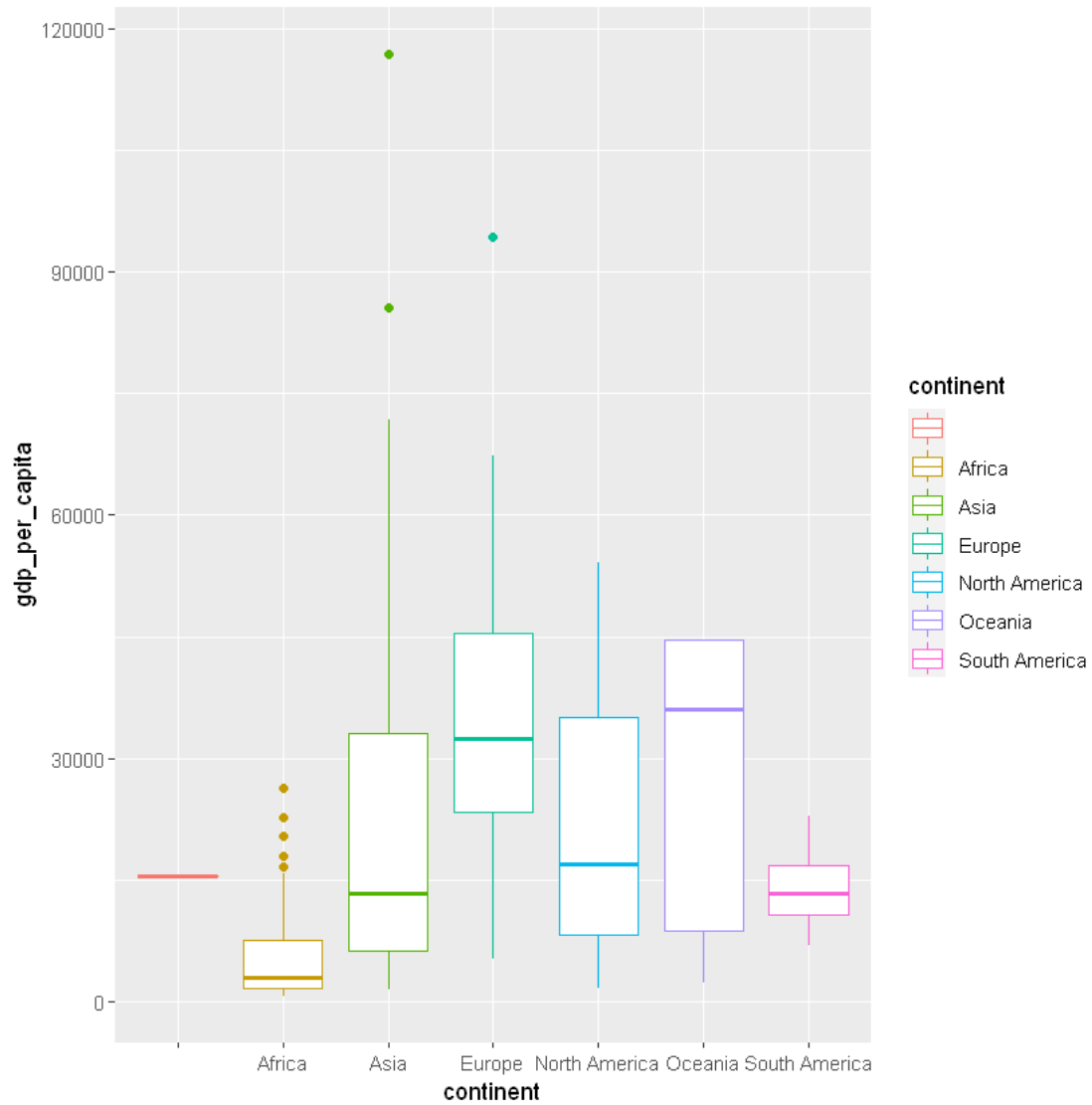
Warning message:

"Removed 6737 rows containing non-finite values (stat\_boxplot)."

Saving 6.67 x 6.67 in image

Warning message:

"Removed 6737 rows containing non-finite values (stat\_boxplot)."



[ ]: