

Lösungen zu den Aufgaben

1. Aufgabe

Laden Sie das Paket `gapminder` (ggf. vorab installieren) und laden Sie daraus den Datensatz `gapminder`. [Hier](#) finden Sie weitere Informationen zum Projekt *Gapminder*.

Tip: Mit `help(gapminder)` bekommen Sie mehr Informationen zum Datensatz.

Hinweis: Beziehen Sie sich auf den Stoff dieses Buches: Ismay, C., & Kim, A. (2019). *ModernDive—An Introduction to Statistical and Data Sciences via R*. <http://moderndive.com/>.

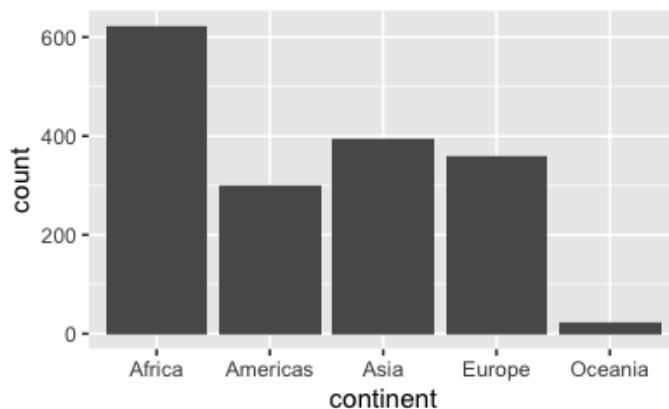
- Visualisieren Sie die Anzahl der Länder pro Kontinent mit einem geeigneten Diagramm!
- Weisen Sie der Füllfarbe die Werte der Variablen zum Kontinent zu.
- Ergänzen Sie `+ scale_fill_viridis_d()`. Was ändert sich?

Lösung

A.

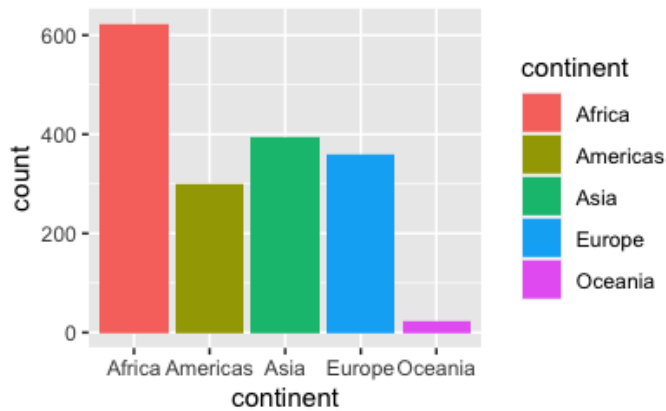
```
library(gapminder) # falls noch nicht geladen
library(tidyverse) # falls noch nicht geladen
data("gapminder") # falls noch nicht geladen
```

```
gapminder %>%
  ggplot() +
  aes(x = continent) +
  geom_bar()
```



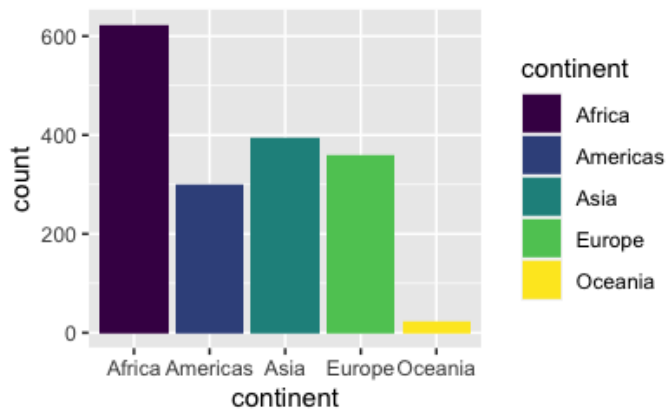
B.

```
gapminder %>%
  ggplot() +
  aes(x = continent, fill = continent) +
  geom_bar()
```



C.

```
gapminder %>%
  ggplot() +
  aes(x = continent, fill = continent) +
  geom_bar() +
  scale_fill_viridis_d()
```



2. Aufgabe

Laden Sie das Paket `gapminder` (ggf. vorab installieren) und laden Sie daraus den Datensatz `gapminder`. [Hier](#) finden Sie weitere Informationen zum Projekt *Gapminder*.

Tip: Mit `help(gapminder)` bekommen Sie mehr Informationen zum Datensatz.

- Erstellen Sie einen Boxplot für jedes *Land* im Datensatz, um die Verteilung der Lebenserwartung zu visualisieren.
- Erstellen Sie einen Boxplot für jedes *Kontinent* im Datensatz, um die Verteilung der Lebenserwartung zu visualisieren.
- Erstellen Sie einen Boxplot für jedes *Kontinent* im Datensatz, um die Verteilung der Lebenserwartung zu visualisieren. Dieses Mal sollen Sie aber vorab den Datensatz zusammenfassen, so dass ein (zeilen-)reduzierter Datensatz entsteht, der für jeden Kontinent eine Zeile umfasst. Wie sinnvoll ist dieses Vorgehen?
- Erstellen Sie einen Boxplot pro Kontinent und weisen Sie der Füllfarbe die Variable zum Kontinent zu.

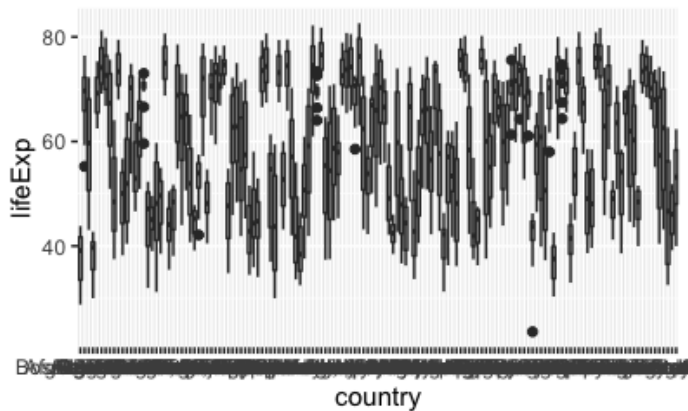
Hinweis: Beziehen Sie sich auf den Stoff dieses Buches: Ismay, C., & Kim, A. (2019). *ModernDive—An Introduction to Statistical and Data Sciences via R*. <http://moderndive.com/>.

Lösung

A.

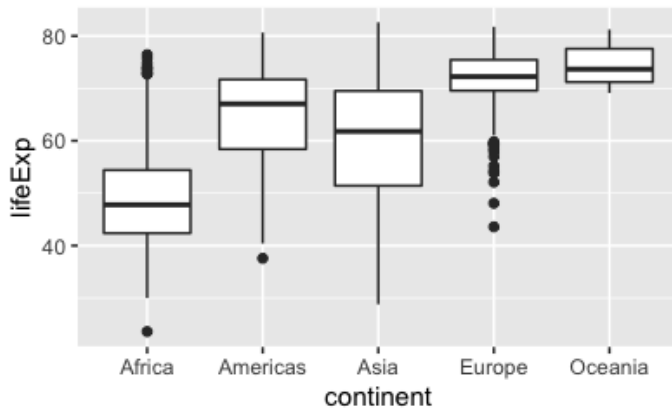
```
library(gapminder) # falls noch nicht geladen
library(tidyverse) # falls noch nicht geladen
data("gapminder") # falls noch nicht geladen
```

```
gapminder %>%
  ggplot() +
  aes(x = country, y = lifeExp) +
  geom_boxplot()
```



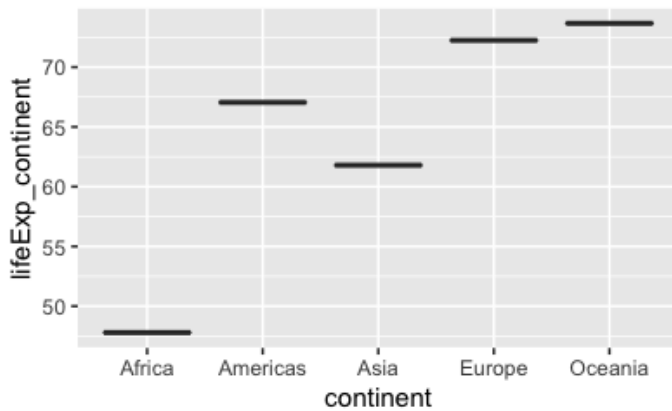
B.

```
gapminder %>%
  ggplot() +
  aes(x = continent, y = lifeExp) +
  geom_boxplot()
```



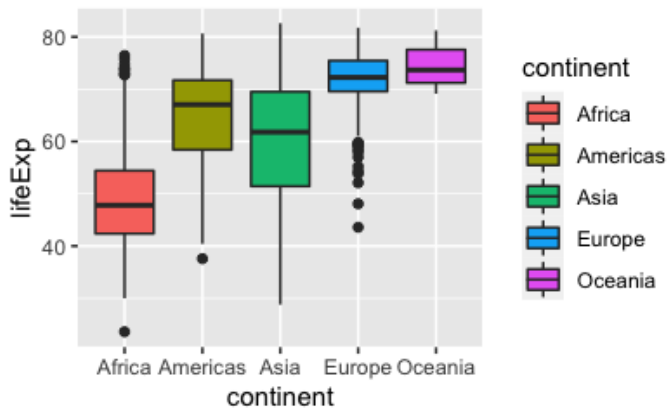
C.

```
gapminder %>%
  group_by(continent) %>%
  summarise(lifeExp_continent = median(lifeExp)) %>%
  ggplot() +
  aes(x = continent, y = lifeExp_continent) +
  geom_boxplot()
```



D.

```
gapminder %>%
  ggplot() +
  aes(x = continent, y = lifeExp, fill = continent) +
  geom_boxplot()
```



Siehe Antworten. Es erfolgt keine automatische

Bewertung dieser Aufgabe.

3. Aufgabe

Laden Sie das Paket `gapminder` (ggf. vorab installieren) und laden Sie daraus den Datensatz `gapminder`. [Hier](#) finden Sie weitere Informationen zum Projekt *Gapminder*.

- Erstellen Sie ein Histogramm, um die Verteilung der Lebenserwartung zu skizzieren.
- Teilen Sie das Histogramm in mehrere Facetten auf, entsprechend der Kontinente.
- Erstellen Sie nun wieder ein Histogramm, aber fügen Sie bei `aes()` noch hinzu, dass die Füllfarbe zum Kontinent zugeordnet werden soll.
- Ersetzen Sie im letzten Diagramm das Geom "Histogramm" durch das Geom "density" (sog. "Dichtediagramm", also `geom_density`). Reduzieren Sie das Alpha der Füllung auf 50%. Welche Variante (Histogramm oder Dichtediagramm) ist sinnvoller, bzw. wann sinnvoller? Warum?

Tip: Mit `help(gapminder)` bekommen Sie mehr Informationen zum Datensatz.

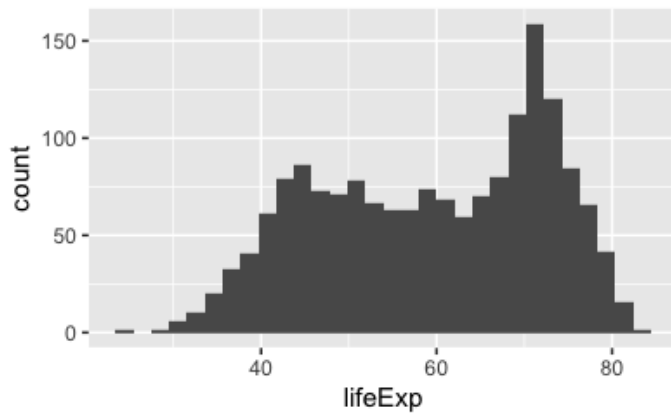
Hinweis: Beziehen Sie sich auf den Stoff dieses Buches: Ismay, C., & Kim, A. (2019). *ModernDive—An Introduction to Statistical and Data Sciences via R*. <http://moderndive.com/>.

Lösung

A.

```
library(gapminder) # falls noch nicht geladen
library(tidyverse) # falls noch nicht geladen
data("gapminder") # falls noch nicht geladen
```

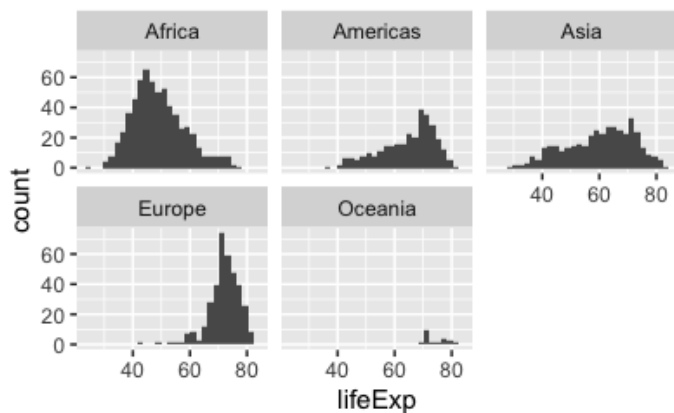
```
gapminder %>%
  ggplot() +
  aes(x = lifeExp) +
  geom_histogram()
```



B.

```
gapminder %>%
  ggplot() +
  aes(x = lifeExp) +
  geom_histogram() +
  facet_wrap(~ continent)
```

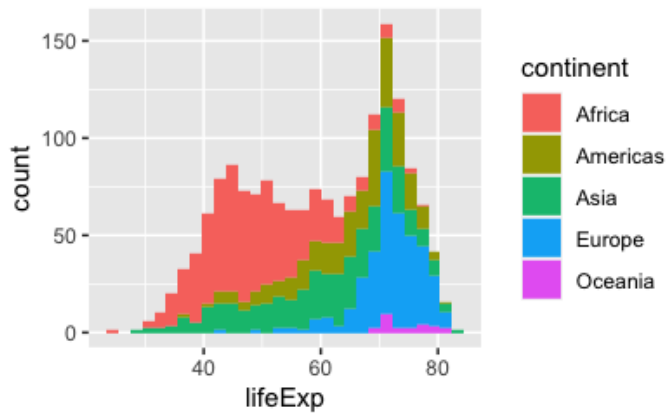
```
## `stat_bin()` using `bins = 30`. Pick better value with
## `binwidth`.
```



C.

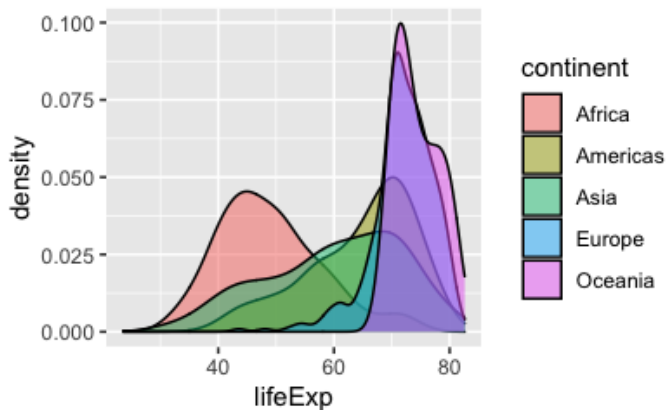
```
gapminder %>%
  ggplot() +
  aes(x = lifeExp, fill = continent) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with
## `binwidth`.
```



D.

```
gapminder %>%
  ggplot() +
  aes(x = lifeExp, fill = continent) +
  geom_density(alpha = .5)
```



Siehe Antworten. Es erfolgt keine automatische Bewertung dieser Aufgabe.

4. Aufgabe

Laden Sie das Paket `gapminder` (ggf. vorab installieren) und laden Sie daraus den Datensatz `gapminder`. [Hier](#) finden Sie weitere Informationen zum Projekt *Gapminder*.

A. Erstellen Sie ein Streudiagramm, das den Zusammenhang von Bruttonationalprodukt und Lebenserwartung widerspiegelt.

Tip: Mit `help(gapminder)` bekommen Sie mehr Informationen zum Datensatz.

B. Logarithmieren Sie die Variable zum Bruttonationalprodukt und erstellen Sie auf dieser Basis das Diagramm erneut.

C. Interpretieren Sie dieses (zweite) Diagramm.

D. Die Punkte im Streudiagramm sind stark überlagert. Wie kann man diese "Overplotting" verringern?

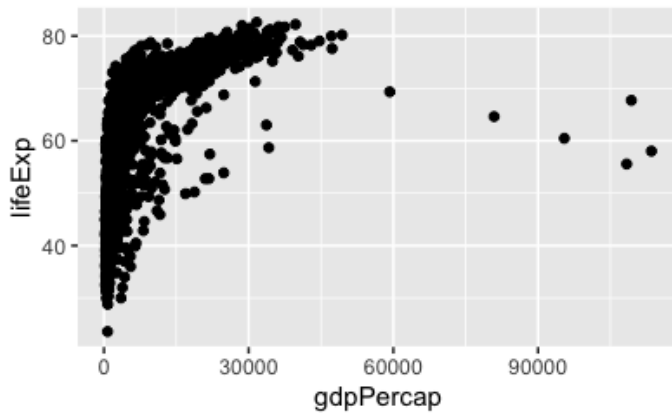
Hinweis: Beziehen Sie sich auf den Stoff dieses Buches: Ismay, C., & Kim, A. (2019). *ModernDive—An Introduction to Statistical and Data Sciences via R*. <http://moderndive.com/>.

Lösung

A.

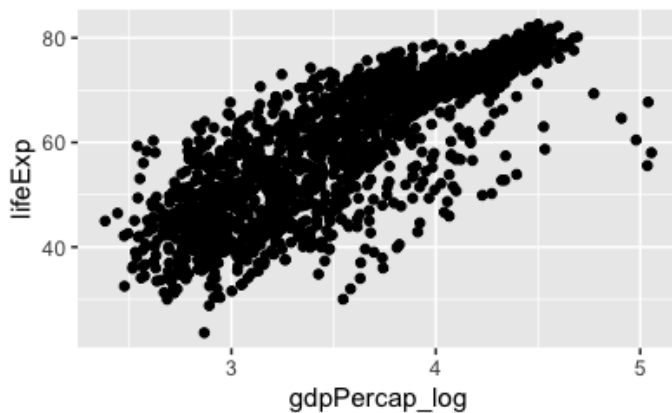
```
library(gapminder)
library(tidyverse)
data("gapminder")

gapminder %>%
  ggplot() +
  aes(x = gdpPercap, y = lifeExp) +
  geom_point()
```



B.

```
gapminder %>%
  mutate(gdpPercap_log = log(gdpPercap, base = 10)) %>%
  ggplot() +
  aes(x = gdpPercap_log, y = lifeExp) +
  geom_point()
```



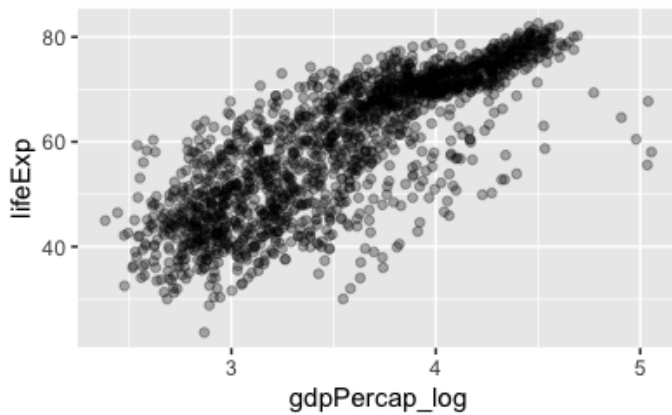
C.

Es findet sich ein linearer, positiver, substanzieller Zusammenhang. Die logarithmierte Variable zeigt die "Nuller" hinter der führenden Eins, also die Größenordnung des Bruttosozialprodukts.

D.

Eine Möglichkeit besteht darin, die Punkte nicht stark zu färben; das "Alpha" der Füllung also zu verringern.

```
gapminder %>%
  mutate(gdpPercap_log = log(gdpPercap, base = 10)) %>%
  ggplot() +
  aes(x = gdpPercap_log, y = lifeExp) +
  geom_point(alpha = .3)
```



Siehe Antworten. Es erfolgt keine automatische Bewertung dieser Aufgabe.

5. Aufgabe

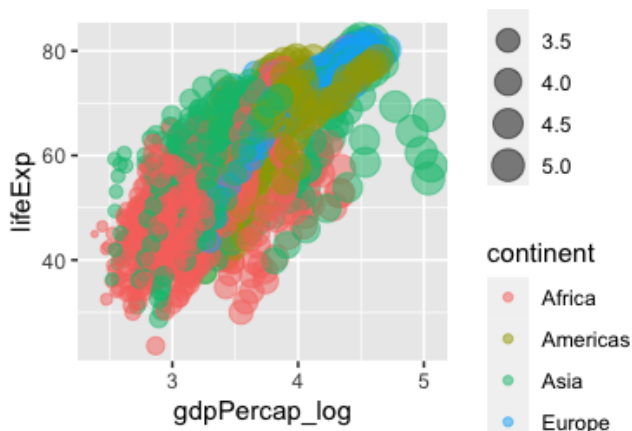
Laden Sie das Paket `gapminder` (ggf. vorab installieren) und laden Sie daraus den Datensatz `gapminder`. [Hier](#) finden Sie weitere Informationen zum Projekt *Gapminder*.

Bauen Sie das folgende Streudiagramm nach!

Tip: Mit `help(gapminder)` bekommen Sie mehr Informationen zum Datensatz.

Hinweis: Beziehen Sie sich auf den Stoff dieses Buches: Ismay, C., & Kim, A. (2019). *ModernDive—An Introduction to Statistical and Data Sciences via R*. <http://moderndive.com/>.

Lösung

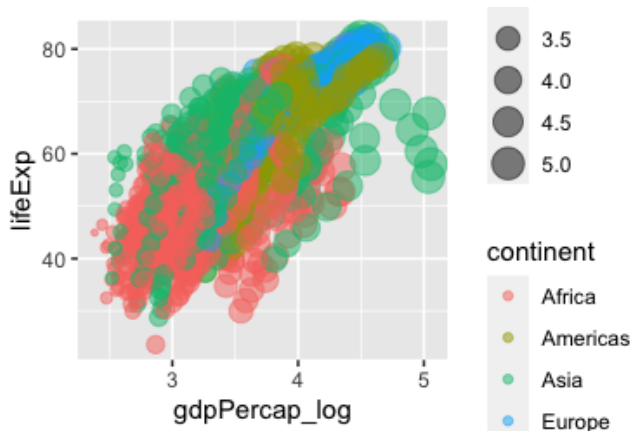


1. Vorbereitung

```
library(gapminder)
library(tidyverse)
data("gapminder")
```

2. Diagramm erstellen

```
gapminder %>%
  mutate(gdpPercap_log = log(gdpPercap, base = 10)) %>%
  ggplot() +
  aes(x = gdpPercap_log, y = lifeExp,
       color = continent,
       size = gdpPercap_log) +
  geom_point(alpha = .5)
```

Siehe Antworten. Es erfolgt keine automatische Bewertung dieser Aufgabe.

6. Aufgabe

In dieser Aufgaben analysieren wir den Datensatz `diamonds`, welcher Merkmale von Diamanten (wie Preis, Schliffart etc.) auflistet. Hier ist ein Blick in den Datensatz:

```
## Rows: 53,940
## Columns: 10
## $ carat <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, ...
## $ cut <ord> Ideal, Premium, Good, Premium, Good, Very Good, ...
## $ color <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, ...
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, V...
## $ depth <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, ...
## $ table <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, ...
## $ price <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338...
## $ x <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, ...
## $ y <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, ...
## $ z <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, ...
```

`diamonds` ist Teil von Tidyverse (genauer gesagt dem Paket `ggplot2`).

Hier ist ein Überblick über die deskriptiven, univariaten Statistiken:

```
Data summary
Name                diamonds
Number of rows      53940
Number of columns   10

Column type frequency:
factor              3
numeric             7

Group variables     None
```

Variable type: factor

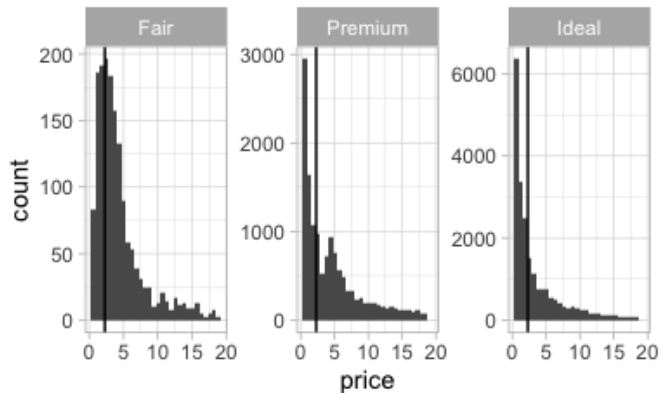
```
skim_variable n_missing complete_rate ordered n_unique top_counts
cut           0             1 TRUE      5 Ide: 21551, Pre: 13791, Ver: 12082, Goo: 4906
color         0             1 TRUE      7 G: 11292, E: 9797, F: 9542, H: 8304
clarity       0             1 TRUE      8 SI1: 13065, VS2: 12258, SI2: 9194, VS1: 8171
```

Variable type: numeric

```
skim_variable n_missing complete_rate mean sd p0 p25 p50 p75 p100 hist
```

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
carat	0	1	0.80	0.47	0.2	0.40	0.70	1.04	5.01	
depth	0	1	61.75	1.43	43.0	61.00	61.80	62.50	79.00	
table	0	1	57.46	2.23	43.0	56.00	57.00	59.00	95.00	
price	0	1	3932.80	3989.44	326.0	950.00	2401.00	5324.25	18823.00	
x	0	1	5.73	1.12	0.0	4.71	5.70	6.54	10.74	
y	0	1	5.73	1.14	0.0	4.72	5.71	6.54	58.90	
z	0	1	3.54	0.71	0.0	2.91	3.53	4.04	31.80	

Betrachten Sie das Histogramm; welcher R-Code hat es erzeugt?



Für alle Diagramme gilt:

A

```
diamonds %>%
  mutate(price = price/1000 %>% round) %>%
  ggplot() +
  aes(x = price) +
  geom_histogram() +
  geom_vline(aes(xintercept = mean(price))) +
  theme_light() +
  facet_wrap(~ cut, scales = "free")
```

B

```
diamonds %>%
  drop_na(cut, price) %>%
  filter(cut %in% c("Fair", "Premium", "Ideal")) %>%
  mutate(price = price/1000 %>% round) %>%
  ggplot() +
  aes(x = price) +
  geom_histogram() +
  geom_vline(aes(xintercept = median(price))) +
  theme_light() +
  facet_wrap(~ cut, scales = "free")
```

C

```
diamonds %>%
  drop_na(cut, price) %>%
  filter(cut %in% c("Fair", "Premium", "Ideal")) %>%
  ggplot() +
  aes(x = price) +
  geom_histogram() +
  geom_vline(aes(xintercept = median(price)))
  facet_wrap(cut)
```

D

```

diamonds %>%
  drop_na(cut, price) %>%
  filter(cut %in% c("Fair", "Premium", "Ideal")) %>%
  ggplot() +
  aes(x = price) +
  geom_histogram() +
  geom_vline(aes(xintercept = median(price))) +
  theme_light() +
  facet_wrap(~ cut, scales = "free")

```

- Code-Beispiel C
- Code-Beispiel D
- Code-Beispiel A
- Code-Beispiel B
- Keine Antwort möglich

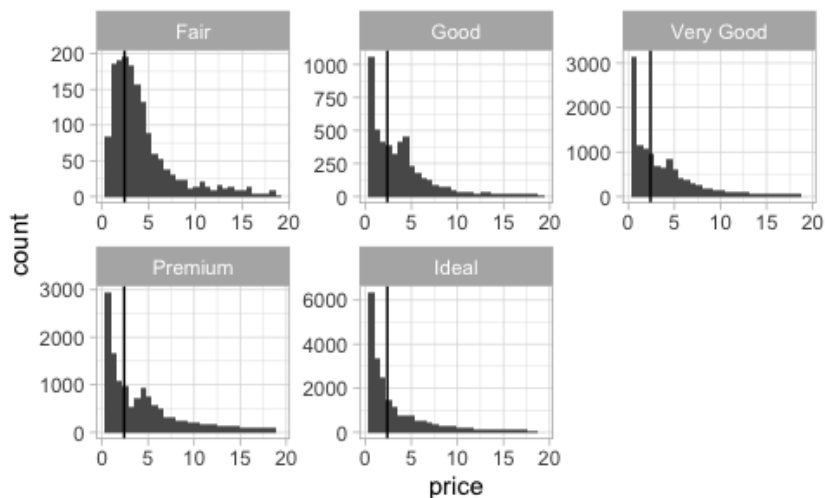
Lösung

Code B passt zum Diagramm.

- Falsch
- Falsch
- Falsch
- Richtig
- Falsch

7. Aufgabe

Welche der folgenden Aussagen zum Diagramm ist korrekt?



- Der vertikale Strich in jedem Bild passt gut zur Position des insgesamten Medians.
- Auf der x-Achse werden Häufigkeiten abgetragen.
- Auf der X-Achse ist eine nominalskalierte Variable abgetragen.
- Die Gruppierungsvariable `cut` wird hier als ordinale Variable, also mit Ordnungsstruktur, verwendet.
- Die Variable `cut` ist eine intervallskalierte Variable.

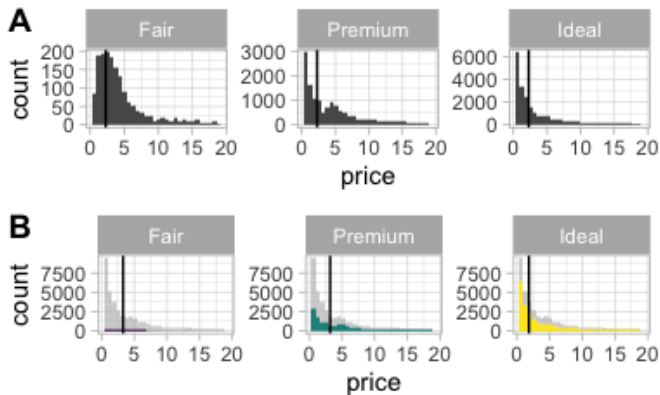
Lösung

Der vertikale Strich in jedem Bild passt jeweils gut zur Position des Medians.

- a. Wahr
- b. Falsch
- c. Falsch
- d. Falsch
- e. Falsch

8. Aufgabe

Betrachten Sie die beiden Diagramm, *A* und *B*; sie zeigen die Verteilung des Preises von Diamanten in Abhängigkeit der Schliffart. Der vertikale Strich zeigt ein Maß der zentralen Tendenz. Welche der Aussagen ist korrekt?



- a. Der vertikale Strich passt *nicht* auf den Median.
- b. Es ist nicht sinnvoll, die Gesamtverteilung zusätzlich zur Verteilung pro Gruppe in jeder Facette darzustellen.
- c. Den globalen Median (für den gesamten Datensatz, also über alle Gruppen hinweg) in jeder Facette darzustellen, ist redundant. Daher ist es besser, in jeder Facetten den Median pro Gruppe darzustellen.
- d. Die Verwendung einer Füllfarbe (Diagramm B) ist hier nicht sinnvoll.

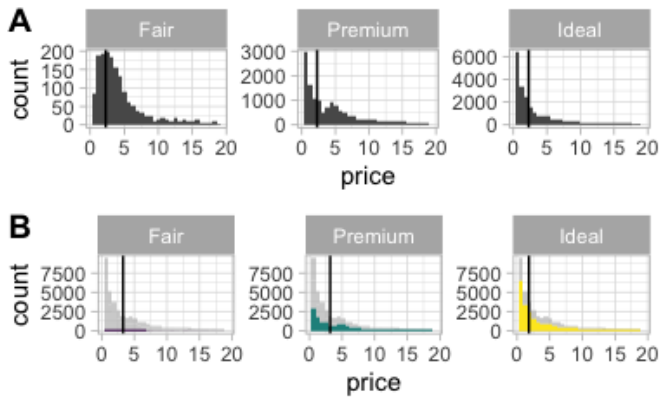
Lösung

Im Diagramm B wird die Gesamtverteilung über die drei Gruppen hinweg (in hellgrau) dargestellt; in den kräftigeren Farbtöne wird die Verteilung pro Gruppe (Schliffart) dargestellt.

- a. Falsch
- b. Falsch
- c. Wahr
- d. Falsch

9. Aufgabe

Betrachten Sie die beiden Diagramm, *A* und *B*; sie zeigen die Verteilung des Preises von Diamanten in Abhängigkeit der Schliffart. Der vertikale Strich zeigt ein Maß der zentralen Tendenz. Welche der Aussagen ist korrekt?



- Im Diagramm A wird ein Maß der zentralen Tendenz gruppenbezogen gezeigt, also den jeweiligen Kennwert der Gruppe (Schliffart) wiedergegeben.
- Insgesamt sind die Verteilung linksschief.
- Im Diagramm B wird die Gesamtverteilung über die drei Gruppen hinweg (in hellgrau) dargestellt; in den kräftigeren Farbtönen wird die Verteilung pro Gruppe (Schliffart) dargestellt.
- Insgesamt sind die Verteilung rechtssteil.

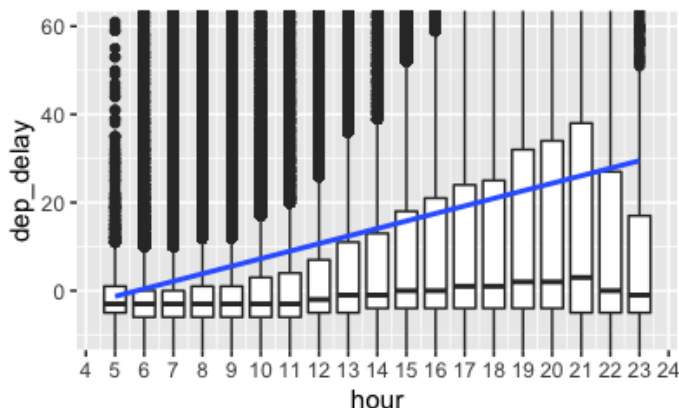
Lösung

Im Diagramm B wird die Gesamtverteilung über die drei Gruppen hinweg (in hellgrau) dargestellt; in den kräftigeren Farbtönen wird die Verteilung pro Gruppe (Schliffart) dargestellt.

- Falsch
- Falsch
- Wahr
- Falsch

10. Aufgabe

Was ist der richtige R-Code, um das Diagramm zu erzeugen?



****R-Code A****

```
flights %>%
  ggplot(aes(x = hour, y = dep_delay)) +
  geom_boxplot(aes(group = hour)) +
  geom_smooth() +
  coord_cartesian(ylim = c(-10, 60)) +
  scale_x_continuous(breaks = 1:24)
```

****R-Code B****

```
flights %>%
  select(arr_delay, hour) %>%
  ggplot(aes(x = hour, y = arr_delay)) +
  geom_boxplot(aes(group = hour)) +
  geom_smooth(method = 'lm') +
  coord_cartesian(ylim = c(-10, 60))
```

****R-Code C****

```
flights %>%
  select(dep_delay, hour) %>%
  ggplot(aes(x = hour, y = dep_delay)) +
  geom_boxplot(aes(group = hour)) +
  geom_smooth(method = 'lm') +
  coord_cartesian(ylim = c(-10, 60)) +
  scale_x_continuous(breaks = 1:24)
```

****R-Code D****

```
flights %>%
  select(dep_delay, hour) %>%
  ggplot(aes(y = hour, x = dep_delay)) +
  geom_boxplot(aes(group = dep_delay)) +
  geom_smooth(method = 'lm') +
  coord_cartesian(ylim = c(-10, 60)) +
  scale_x_continuous(breaks = 1:24)
```

****R-Code E****

```
flights %>%
  select(dep_delay, hour) %>%
  ggplot(aes(x = hour, y = dep_delay)) +
  geom_boxplot(aes(group = hour)) +
  coord_cartesian(ylim = c(-10, 60)) +
  scale_x_continuous(breaks = 1:24)
```

- a. A
- b. B
- c. C
- d. D
- e. E

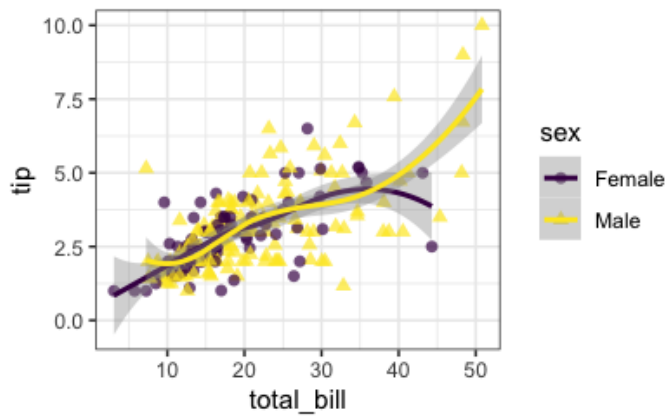
Lösung

Der Code-Auszug c ist korrekt.

- a. Falsch
- b. Falsch
- c. Wahr
- d. Falsch
- e. Falsch

11. Aufgabe

Was ist der richtige R-Code, um das Diagramm zu erzeugen?



****R-Code A****

```
tips %>%
  ggplot(aes(x = total_bill, y = tip,
             color = sex, shape = sex)) +
  geom_point() +
  scale_color_viridis_d() +
  theme_bw()
```

****R-Code B****

```
tips %>%
  ggplot(aes(x = total_bill, y = tip,
             color = sex, shape = sex)) +
  geom_point(size = 2) +
  geom_smooth() +
  scale_color_viridis_d() +
  theme_bw()
```

****R-Code C****

```
tips %>%
  ggplot(aes(y = total_bill, x = tip,
             color = sex, shape = sex)) +
  geom_point(size = 2) +
  geom_smooth() +
  scale_color_viridis_d() +
  theme_bw()
```

****R-Code D****

```
tips %>%
  ggplot(aes(x = total_bill, y = tip,
             color = sex, shape = sex)) +
  geom_point(size = 2) +
  geom_smooth()
```

****R-Code E****

```
tips %>%
  ggplot(aes(x = total_bill, y = tip,
             color = sex)) +
  geom_point(size = 1) +
  geom_smooth() +
  scale_color_viridis_d()
```

- a. A
- b. B
- c. C
- d. D
- e. E

Lösung

Der Code-Auszug B ist korrekt.

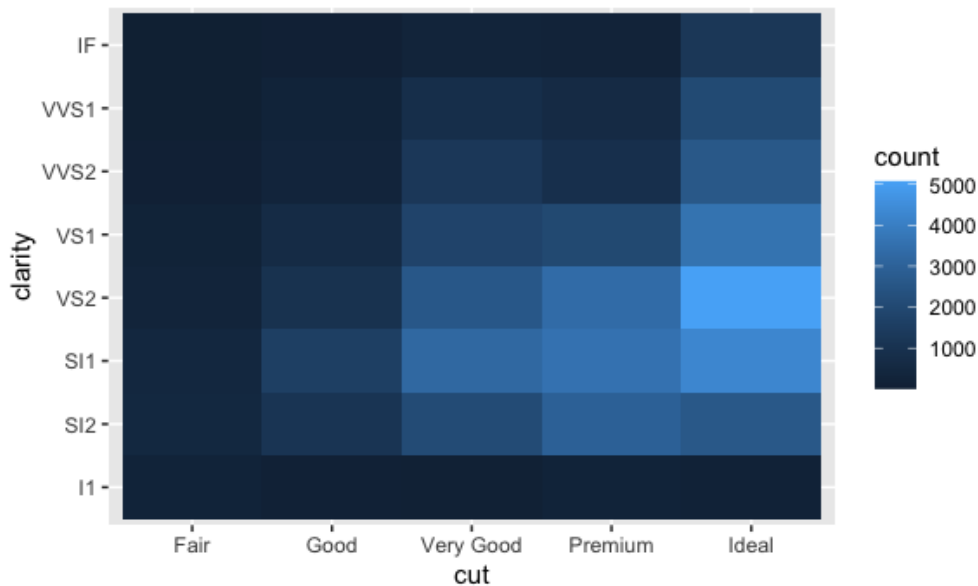
- a. Falsch
- b. Wahr
- c. Falsch
- d. Falsch
- e. Falsch

12. Aufgabe

Je nach Datenanalyse sind verschiedene Arten von Diagramm sinnvoll. Ein Diagrammtyp namens *Heatmap* lässt sich in R z.B. so erstellen:

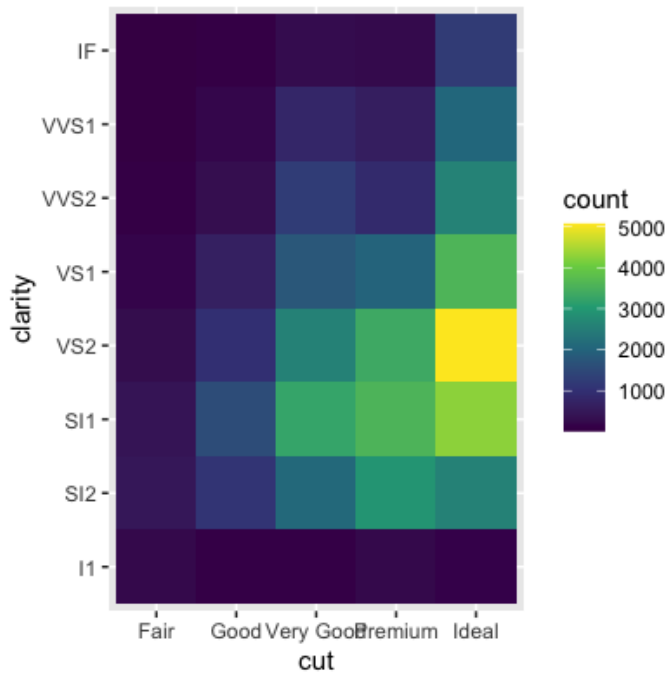
```
library(tidyverse)
data("diamonds")

p1 <-
  ggplot(diamonds) +
    aes(x = cut, y = clarity) +
    geom_bin2d()
p1
```



Ändern wir noch das Farbschema, damit die Farbunterschiede deutlicher zutage treten, wir setzen sozusagen die Skibrille mit den gelben Gläsern auf.

```
p1 +
  scale_fill_viridis_c()
```

plot of chunk unnamed-chunk-3

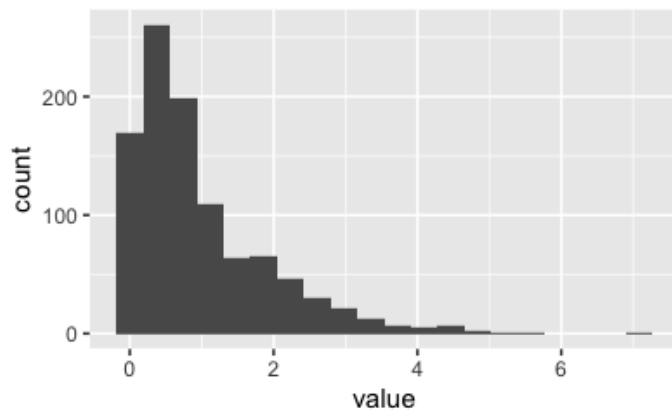
Welche Aussage passt am besten zu diesem Diagramm?

- a. Es handelt sich um eine univariate Häufigkeitsanalyse.
- b. Je dunkler die Farbe, desto häufiger die Kategorie.
- c. Es handelt sich um eine Visualisierung einer Kontingenztabelle.
- d. Es wurden zwei metrische Variablen untersucht.
- e. Heatmaps sind insgesamt eine wenig geeignete Form der Datenanalyse

13. Aufgabe

Betrachten Sie das Histogramm. Welcher Boxplot spiegelt das Histogramm am genauesten wider?

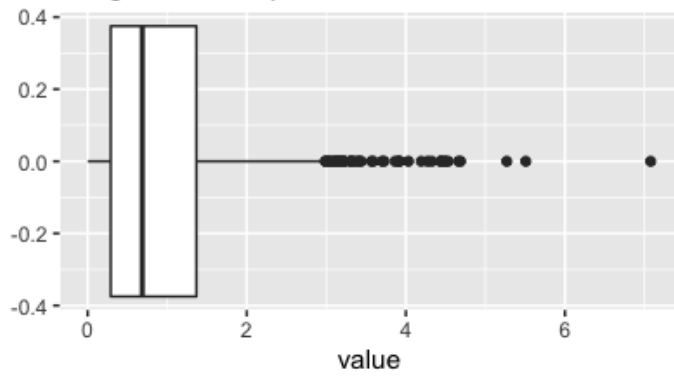
Histogramm:



Boxplots:

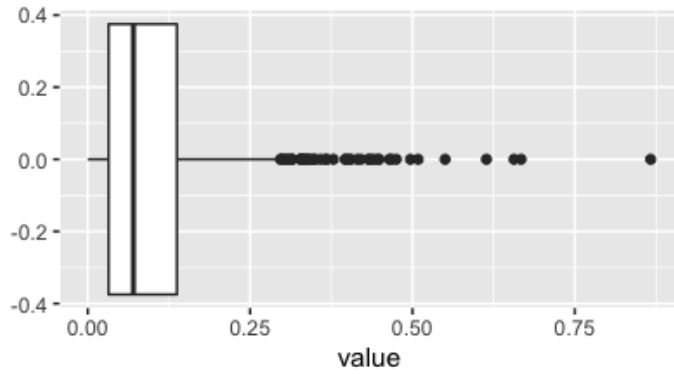
[[1]]

Diagramm Boxplot A



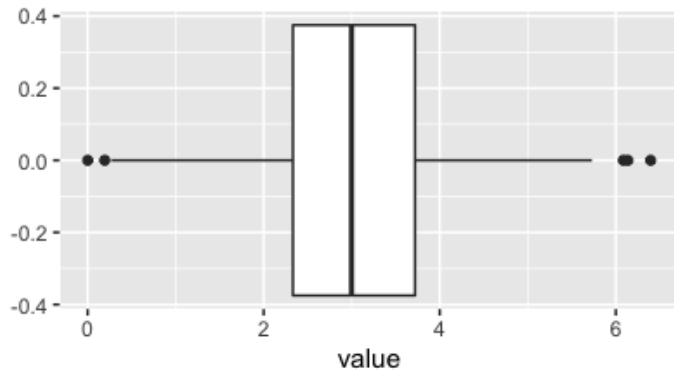
```
##  
## [[2]]
```

Diagramm Boxplot B



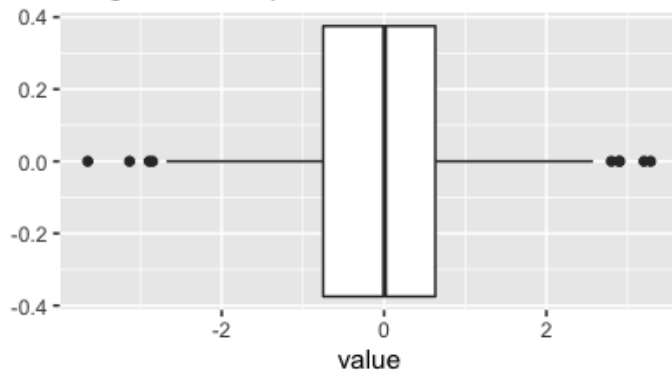
```
##  
## [[3]]
```

Diagramm Boxplot C



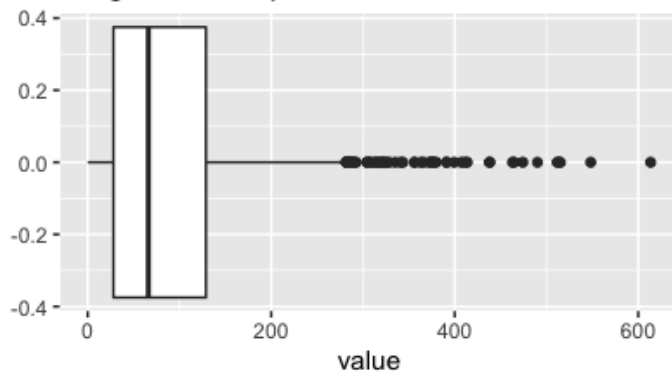
```
##  
## [[4]]
```

Diagramm Boxplot D



```
##  
## [[5]]
```

Diagramm Boxplot E



- a. Boxplot A
- b. Boxplot B
- c. Boxplot C
- d. Boxplot D
- e. Boxplot E

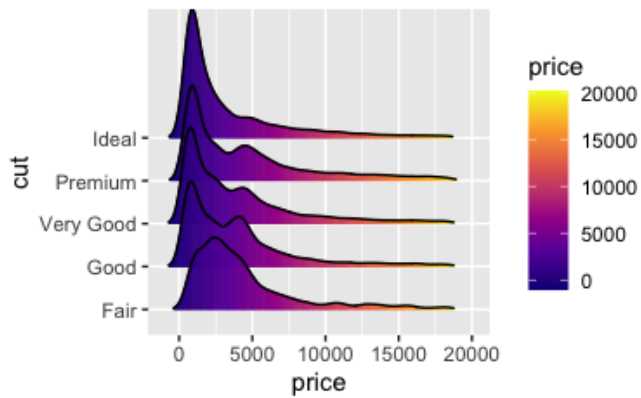
Lösung

Boxplot `Boxplot A` spiegelt das Histogramm am genauesten wider.

- a. Wahr
- b. Falsch
- c. Falsch
- d. Falsch
- e. Falsch

14. Aufgabe

Wir analysieren die Verteilung des Preises (`price`) von Diamanten, gruppiert nach Schliffart (`cut`). Betrachten Sie die Histogramme. Welche Aussage ist korrekt?



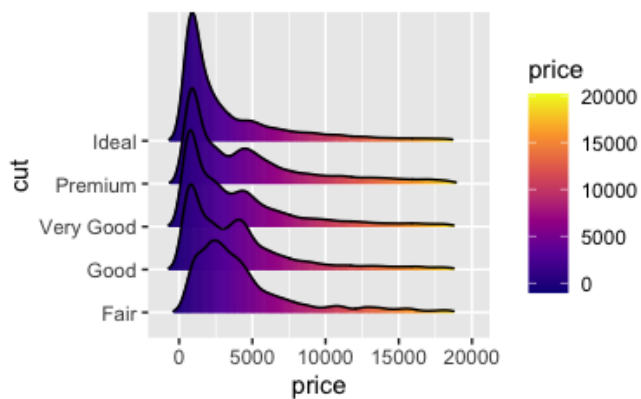
- Die Färbung (Füllfarbe) kodiert die Schliffart (`cut`).
- Die Mittelwerte der Histogramme sind identisch.
- Die Mediane der Histogramme sind identisch.
- Einige Histogramme sind normalverteilt.
- Die Histogramme sind (alle) rechtsschief.

Lösung

- Falsch
- Falsch
- Falsch
- Falsch
- Richtig

15. Aufgabe

Wir analysieren die Verteilung des Preises (`price`) von Diamanten, gruppiert nach Schliffart (`cut`). Betrachten Sie die Histogramme. Welche Aussage ist korrekt?



- Die Schliffart `Premium` ist bimodal verteilt.
- Die Histogramme sind (alle) symmetrisch.
- Die Schliffart `Fair` ist bimodal verteilt.
- Bei allen Schliffarten ist der Median kleiner als der Mittelwert.

Lösung

- Falsch
- Falsch

- c. Falsch
- d. Richtig

16. Aufgabe

Wählen Sie den am besten treffenden Wert des Korrelationskoeffizientens im Streudiagramm.



- a. +.90
- b. 1
- c. -.90
- d. 0
- e. -1

Lösung

$r \approx 0$.

- a. Falsch
- b. Falsch
- c. Falsch
- d. Wahr
- e. Falsch

17. Aufgabe

Welche Aussage zu dieser R-Syntax ist *falsch*:

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy))
```

- a. Der Datensatz heißt `mpg`.
- b. Als Geom werden Punkte angegeben.
- c. Auf der X-Achse soll die Variable `displ` stehen.
- d. Es sollen zwei Plots gezeichnet werden (daher `+`).
- e. Auf der Y-Achse soll die Variable `hwy` stehen.

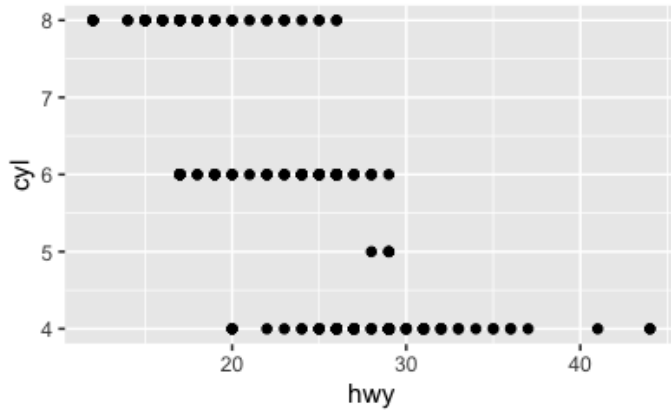
Lösung

Es sollen zwei Plots gezeichnet werden (daher `+`).

- a. Falsch
- b. Falsch
- c. Falsch
- d. Wahr
- e. Falsch

18. Aufgabe

Gesucht ist die Syntax, die folgendes Diagramm erstellt (Datensatz `mpg`):



A

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(y = hwy, x = cyl))
```

B

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = hwy, y = cyl))
```

C

```
data(mpg)
ggplot(data = data) +
  geom_point(aes(x = hwy, y = cyl))
```

D

```
data(mpg)
geom_point(mapping = aes(x = hwy, y = cyl))
```

E

```
data(mpg)
ggplot(data = mpg) +
  geom_point((x = hwy, y = cyl))
```

- a. Antwort A
- b. Antwort B
- c. Antwort C
- d. Antwort D
- e. Antwort E

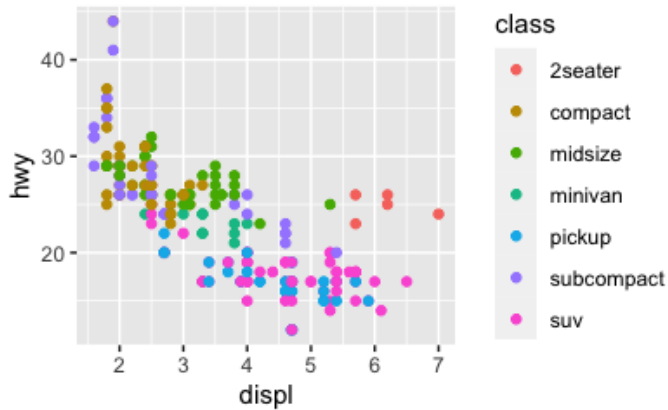
Lösung

Antwort B

- a. Falsch
- b. Wahr
- c. Falsch
- d. Falsch
- e. Falsch

19. Aufgabe

Gesucht ist die Syntax, die folgendes Diagramm erstellt (Datensatz `mpg`):



A

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

B

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy), color = class)
```

C

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = class))
```

D

```
data(mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = class))
```

E

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(color = class), x = displ, y = hwy)
```

- a. Antwort A
- b. Antwort B
- c. Antwort C
- d. Antwort D
- e. Antwort E

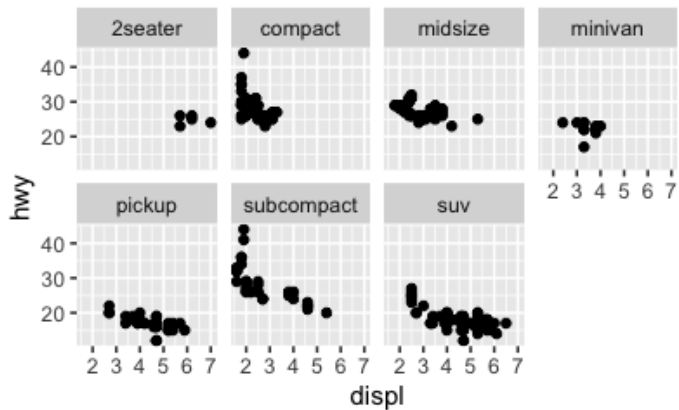
Lösung

Antwort C

- a. Falsch
- b. Falsch
- c. Wahr
- d. Falsch
- e. Falsch

20. Aufgabe

Gesucht ist die Syntax, die folgendes Diagramm erstellt (Datensatz `mpg`):



A

```
data(mpg)
ggplot(data = mpg)
  + geom_point(mapping = aes(x = displ, y = hwy))
  + facet_wrap(~ class, nrow = 2)
```

B

```
data(mpg)
ggplot(data = mpg) +
  geom_point(x = displ, y = hwy) +
  facet_wrap(~ class, nrow = 2)
```

C

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping(x = displ, y = hwy) +
  facet_wrap(~ class, nrow = 2)
```

D

```
data(mpg) +
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ; y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```

E

```
data(mpg)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```

- a. Antwort A
- b. Antwort B

- c. Antwort C
- d. Antwort D
- e. Antwort E

Lösung

Antwort E

- a. Falsch
- b. Falsch
- c. Falsch
- d. Falsch
- e. Wahr

21. Aufgabe

Welche der folgenden Erweiterungen für `ggplot2` gibt es *nicht*?

- a. Anatomiediagramme
- b. Animationen
- c. Diagnostiken zur Überprüfung der Modellannahmen von Regressionsmodellen
- d. Halbierung von Geomen (wer braucht sowas?)
- e. Fortgeschrittene Tortendiagramme
- f. 3D

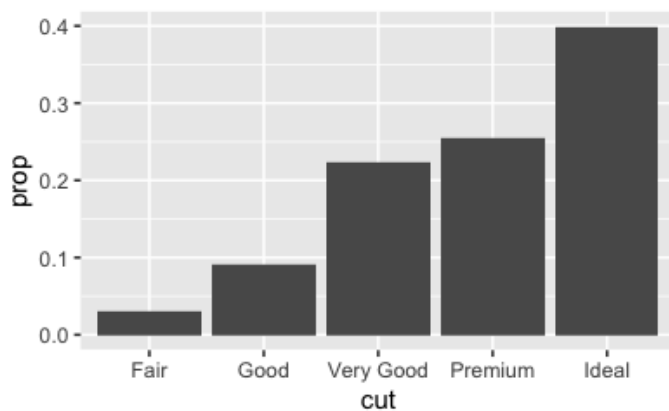
Lösung

Diagramm E

- a. [Falsch](#), s. [hier](#)
- b. [Falsch](#)
- c. [Falsch](#)
- d. [Falsch](#)
- e. [Wahr ja, wirklich](#)
- f. [Falsch](#)

22. Aufgabe

Gesucht ist die Syntax, die folgendes Diagramm erstellt (Datensatz `diamonds`):



A

```
data(diamonds)
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut, y = prop, group = 1))
```

B

```
data(diamonds)
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut, y = prop))
```

C

```
data(diamonds)
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut, y = stat(prop), group = 1))
```

D

```
data(mpg) +
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut, y = stat(prop)), group = 1)
```

E

```
data(mpg)
ggplot(data = diamonds)
+ geom_bar(mapping == aes(x = cut, y = stat(prop), group = 1))
```

- a. Antwort A
- b. Antwort B
- c. Antwort C
- d. Antwort D
- e. Antwort E

Lösung

Antwort C

- a. Falsch
- b. Falsch
- c. Wahr
- d. Falsch
- e. Falsch

23. Aufgabe

Welche Aussage zu `stat_summary()` (bei `ggplot2`) ist korrekt?

- a. summarises the x values for each y value, to draw attention to the summary that you're computing
- b. summarises the unique x values for each unique y value, to draw attention to the summary that you're computing
- c. summarises the x values
- d. summarises the x values for each unique y value, to draw attention to the summary that you're computing
- e. summarises the y values for each unique x value, to draw attention to the summary that you're computing

Lösung

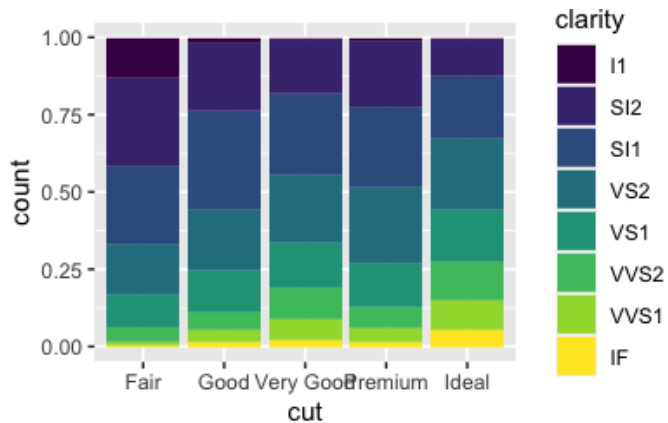
summarises the y values for each unique x value, to draw attention to the summary that you're computing

vgl. [hier](#)

- a. Falsch
- b. Falsch
- c. Falsch
- d. Falsch
- e. Wahr

24. Aufgabe

Gesucht ist die Syntax, die folgendes Diagramm erstellt (Datensatz `diamonds`):



A

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity),)
```

B

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity), position = "fill")
```

C

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity), position = "dodge")
```

D

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity), position = "identity")
```

E

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity, position = "fill"))
```

- a. Antwort A
- b. Antwort B
- c. Antwort C
- d. Antwort D
- e. Antwort E

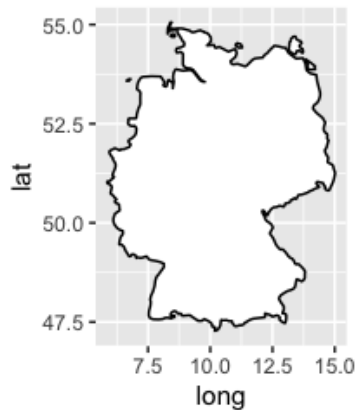
Lösung

Antwort B

- a. Falsch
- b. Wahr
- c. Falsch
- d. Falsch
- e. Falsch

25. Aufgabe

Gesucht ist die Syntax, die folgendes Diagramm erstellt:



A

```
de <- map_data("world", region = "Germany")  
  
ggplot(de, aes(long, lat, group = group)) +  
  geom_polygon(fill = "white", colour = "black")
```

B

```
de <- map_data("world")  
  
ggplot(de, aes(long, lat, group = group)) +  
  geom_polygon(fill = "white", colour = "black")
```

C

```
de <- map_data("world")  
  
ggplot(de, aes(long, lat, group = group)) +  
  geom_polygon(fill = "white", colour = "black") +  
  coord_quickmap()
```

D

```
de <- map_data("world", region = "Germany")  
  
ggplot(de, aes(long, lat, group = group)) +  
  geom_polygon(fill = "white", colour = "black") +  
  coord_quickmap()
```

E

```
de <- map_data("world", region = "France")  
  
ggplot(de, aes(long, lat, group = group)) +  
  geom_polygon(fill = "white", colour = "black") +  
  coord_quickmap()
```

- a. Antwort A
- b. Antwort B
- c. Antwort C
- d. Antwort D
- e. Antwort E

Lösung

Antwort D; s. [hier](#) eine Vertiefung

- a. Falsch
- b. Falsch
- c. Falsch
- d. Wahr
- e. Falsch

26. Aufgabe

`ggplot2` besteht aus mehreren “Bausteinen” oder “Schichten”, die zusammen kombiniert werden können, und so ein Diagramm erstellen.

Welches der folgenden “Schichten” ist *nicht* Teil von `ggplot2`?

- a. DATA
- b. GEOM_FUNCTION
- c. FACET_FUNCTION
- d. VARIABLES
- e. POSITION

Lösung

vgl. [hier](#)

```
`ggplot(data = <DATA>) + <GEOM_FUNCTION>( mapping = aes(<MAPPINGS>), stat = <STAT>, position =  
<POSITION> ) + <COORDINATE_FUNCTION> + <FACET_FUNCTION>
```

- a. Falsch
- b. Falsch
- c. Falsch
- d. Wahr
- e. Falsch