

Explorative Datenanalyse des S&P 500

Valentín Schwarz

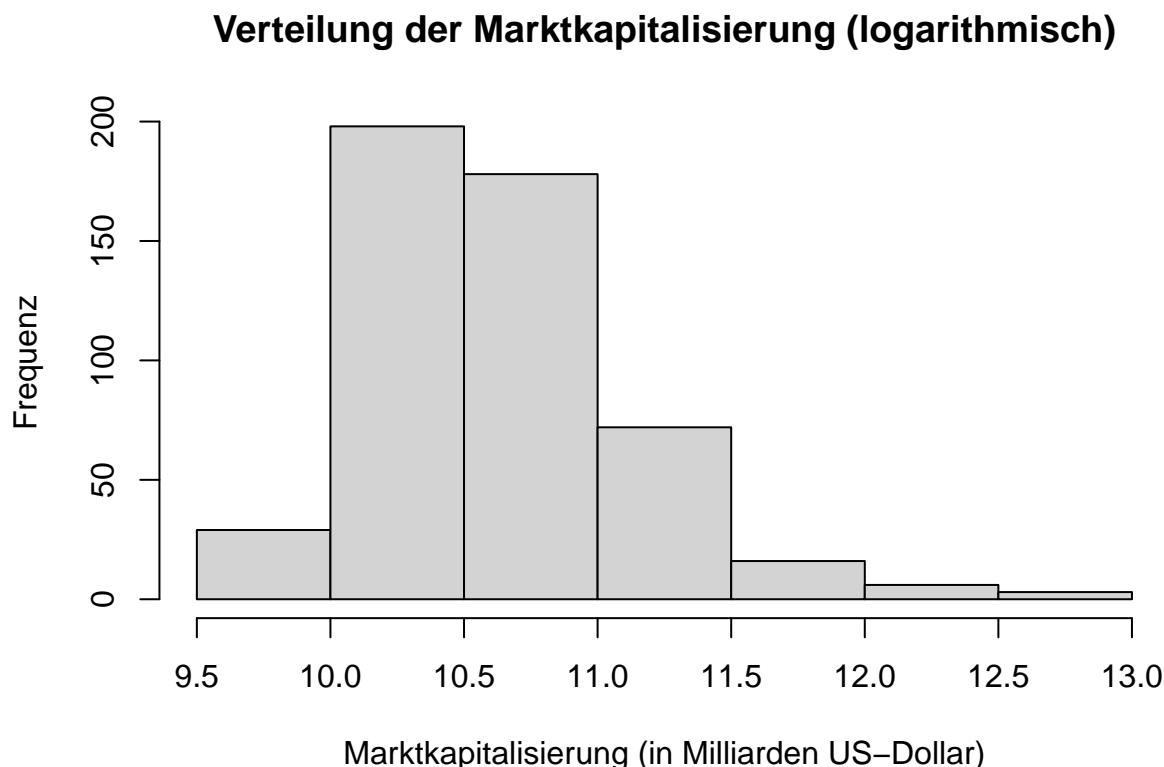
19. Dezember 2025

Aufgabe 1

Diese Grafik zeigt, welche Marktkapitalisierungen am häufigsten vorkommen. Für eine solche Analyse eignet sich ein Histogramm besonders gut. Beim ersten Plot wurde jedoch deutlich, dass eine lineare Darstellung der x-Achse wenig aussagekräftig ist, da die Marktkapitalisierungen eine grosse Spannweite aufweisen. Daher wurde die x-Achse logarithmiert, um die Verteilung klarer und besser interpretierbar darzustellen.

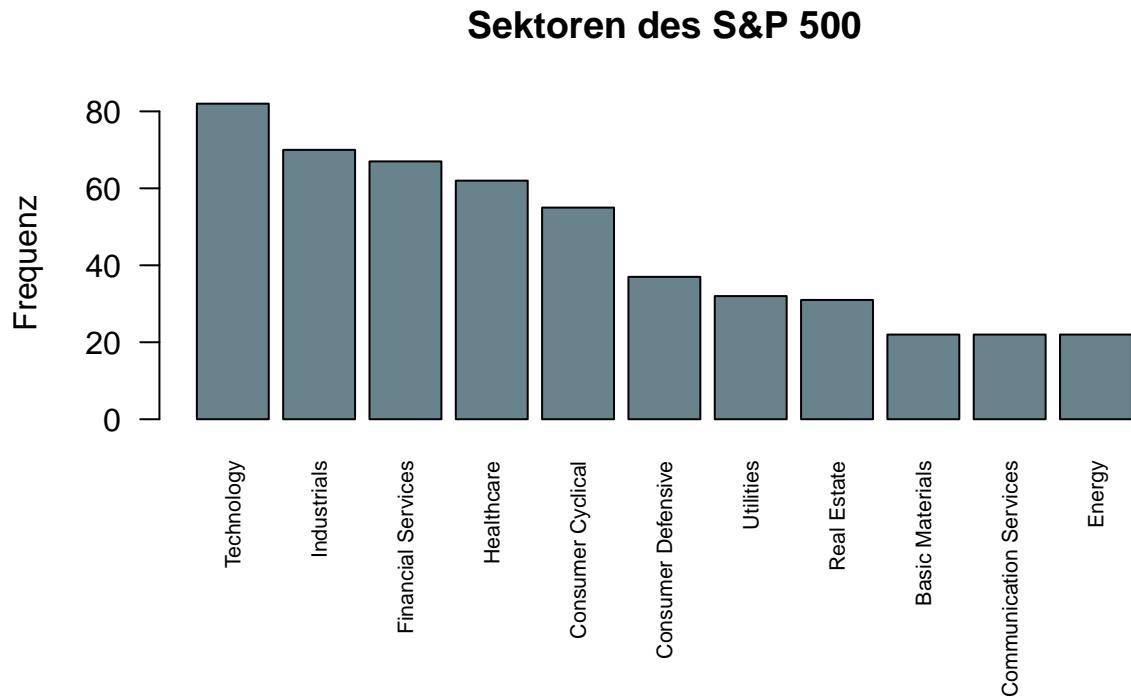
```
data <- read.csv("sp500_companies.csv")

hist(log10(data$Marketcap),
     xlab = "Marktkapitalisierung (in Milliarden US-Dollar)",
     ylab = "Frequenz",
     main = "Verteilung der Marktkapitalisierung (logarithmisch)")
```



Bei der zweiten Grafik haben ich ein Balkendiagramm gewählt, um die Sektoren des S&P 500 nach der Anzahl der Unternehmen zu visualisieren. Die Sektoren wurden dabei absteigend nach Grösse sortiert, sodass der größte Sektor (Technology) zuerst erscheint. Eine Herausforderung war es, die x-Achsen-Beschriftungen für alle Sektoren so anzusehen, dass sie nicht abgeschnitten werden. Durch Anpassung der Plot-Ränder (`par(mar = c(10, 4, 4, 2) + 0.1)`) und der Schriftgrösse (`cex.names = 0.7`) konnten die Sektorennamen lesbar dargestellt werden.

```
data <- read.csv("sp500_companies.csv")
par(mar = c(10, 4, 4, 2) + 0.1)
barplot(height = sort(table(data$Sector), decreasing = TRUE),
        names.arg = names(data$Sector),
        main = "Sektoren des S&P 500",
        ylab = "Frequenz",
        xlab = "",
        las = 2,
        cex.names = 0.7,
        col = "lightblue4")
```



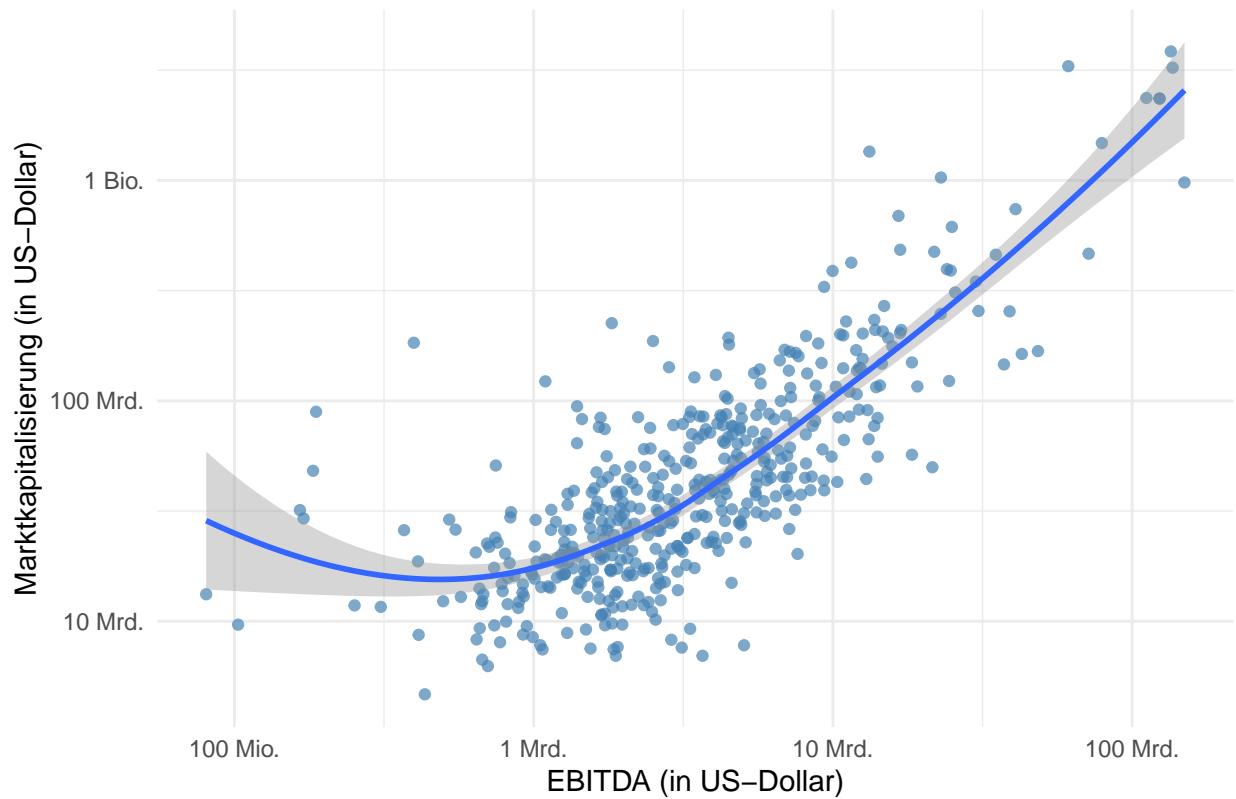
Aufgabe 2

Die Grafik zeigt ein Streudiagramm, das die Beziehung zwischen der Marktkapitalisierung und dem EBITDA darstellt. Aufgrund der sehr grossen Wertebereiche beider Variablen wurde, wie bereits in Aufgabe 1, schnell deutlich, dass eine lineare Darstellung nicht geeignet ist. Daher wurden sowohl die x- als auch die y-Achse mithilfe von `scale_x_log10()` und `scale_y_log10()` logarithmiert, um die Punkteverteilung sichtbar zu machen. Die verwendeten breaks und labels wurden anschliessend manuell angepasst, damit die Achsen trotz der logarithmischen Skalierung gut lesbar bleiben und die Größenordnungen klar erkennbar sind. Durch die Visualisierung lässt sich eine positive Korrelation beobachten: Unternehmen mit hoher Marktkapitalisierung weisen in der Regel auch ein höheres EBITDA auf. Die hinzugefügte Trendlinie (`geom_smooth()`) unterstützt diesen Zusammenhang zusätzlich.

```
data <- read.csv("sp500_companies.csv")
library(ggplot2)
library(scales)

p <- ggplot(data, aes(x = Ebitda, y = Marketcap)) +
  geom_point(color = "steelblue", alpha = 0.7) +
  scale_x_log10(
    breaks = c(1e6, 1e7, 1e8, 1e9, 1e10, 1e11),
    labels = c("1 Mio.", "10 Mio.", "100 Mio.", "1 Mrd.", "10 Mrd.", "100 Mrd.")
  ) +
  scale_y_log10(
    breaks = c(1e8, 1e9, 1e10, 1e11, 1e12),
    labels = c("100 Mio.", "1 Mrd.", "10 Mrd.", "100 Mrd.", "1 Bio.")
  ) +
  labs(
    title = "Marktkapitalisierung vs. EBITDA (logarithmiert)",
    x = "EBITDA (in US-Dollar)",
    y = "Marktkapitalisierung (in US-Dollar)"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_smooth()
print(p)
```

Marktkapitalisierung vs. EBITDA (logarithmiert)

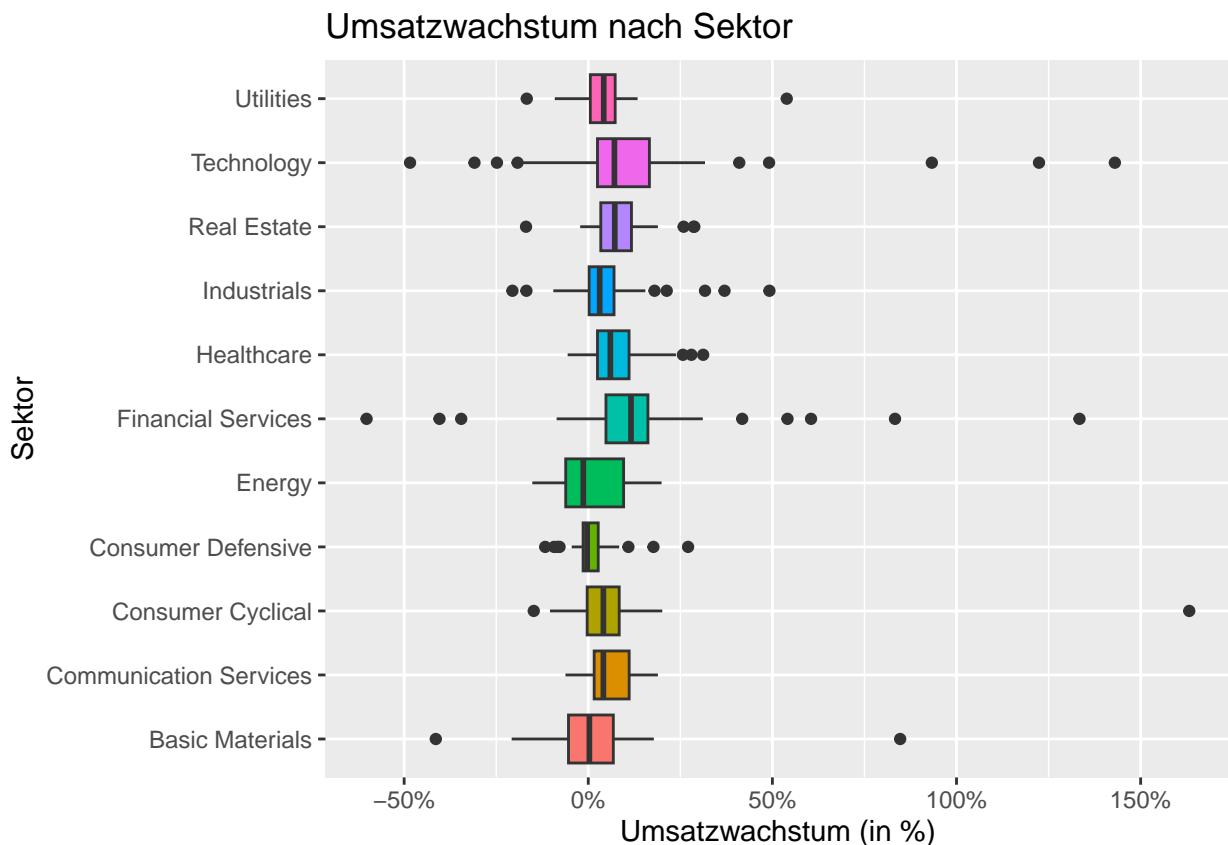


Der Boxplot zeigt das Umsatzwachstum (Revenuegrowth) nach Sektor, wobei die Sektoren auf der y-Achse angeordnet sind. Durch das horizontale Layout (`coord_flip()`) bleibt die Darstellung auch bei vielen Kategorien gut lesbar. Die Wachstumsraten werden als Prozentwerte formatiert, was die Interpretation erleichtert. Die Farbfüllung nach Sektor wurde beibehalten, die Legende hingegen ausgeblendet, da die Sektorbezeichnungen direkt ersichtlich sind.

```
data <- read.csv("sp500_companies.csv")

library(ggplot2)
library(scales)

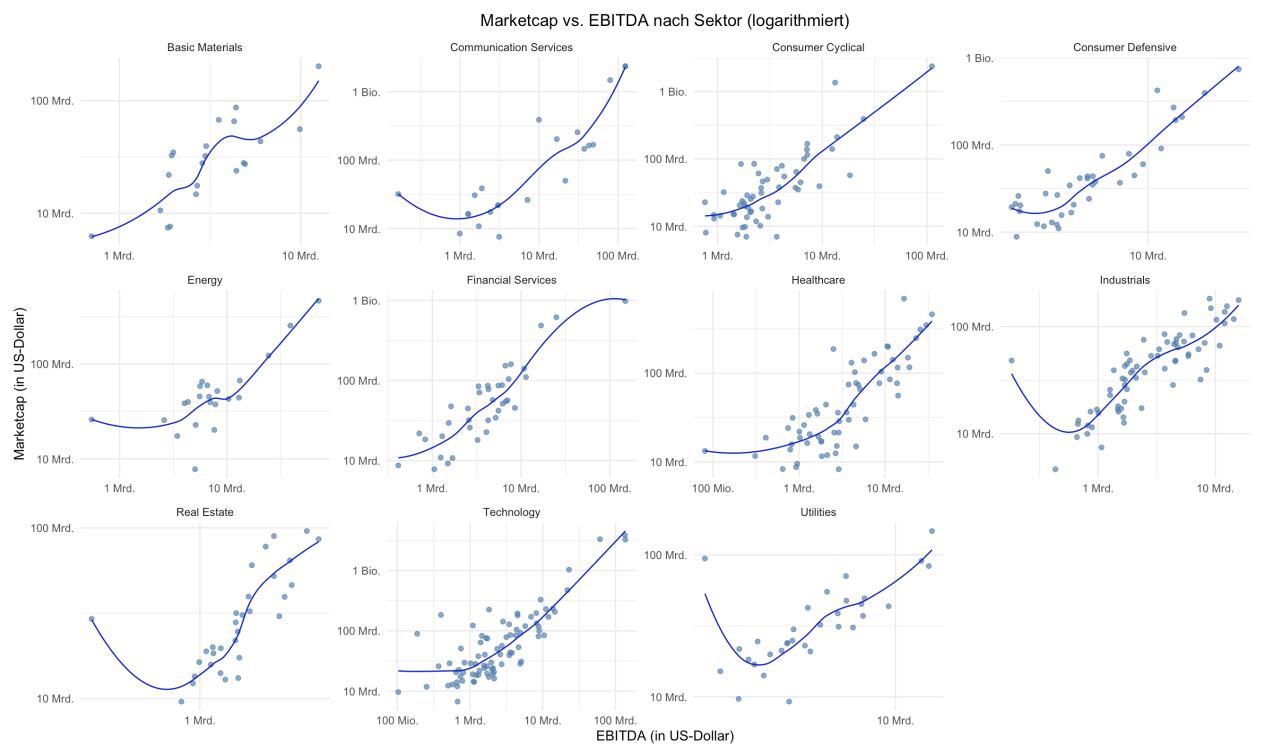
ggplot(data, aes(x = Sector, y = Revenuegrowth, fill = Sector)) +
  geom_boxplot() +
  coord_flip() +
  scale_y_continuous(
    labels = percent_format(accuracy = 1)
  ) +
  labs(
    title = "Umsatzwachstum nach Sektor",
    x = "Sektor",
    y = "Umsatzwachstum (in %)"
  ) +
  theme(legend.position = "none")
```



Aufgabe 4

In dieser Aufgabe haben ich eine komplexere Darstellung erstellt, bei der die Beziehung zwischen Marktkapitalisierung und EBITDA nach Sektor visualisiert wird. Dazu wurde ein Facet-Wrap verwendet, sodass für jeden Sektor ein eigenes Diagramm entsteht und die Skalen individuell angepasst werden (`scales = "free"`). Die Punkte werden in Stahlblau dargestellt (`geom_point`) und leicht transparent gesetzt (`alpha = 0.7`), um Überlappungen sichtbar zu machen. Die x- und y-Achsen sind logarithmiert, um die grosse Spannweite der Werte darzustellen. Die Trendlinie (`geom_smooth()`) wurde hinzugefügt, wobei die Standardabweichung entfernt wurde (`se = FALSE`), um die Darstellung übersichtlicher zu halten. Linienstärke und Darstellung der Trendlinie wurden entsprechend angepasst, damit die Tendenzen innerhalb der einzelnen Sektoren gut erkennbar sind. Die Umsetzung verlief problemlos, und die Visualisierung zeigt klar die Zusammenhänge zwischen Marktkapitalisierung und EBITDA in den einzelnen Sektoren. Ganz nebenbei ist der Plot auch optisch gelungen.

```
v <- ggplot(data, aes(x = Ebitda, y = Marketcap)) +
  geom_point(color = "steelblue", alpha = 0.7) +
  facet_wrap(~ Sector, scales = "free") +
  scale_x_log10(
    breaks = c(1e6, 1e7, 1e8, 1e9, 1e10, 1e11),
    labels = c("1 Mio.", "10 Mio.", "100 Mio.", "1 Mrd.", "10 Mrd.", "100 Mrd.")
  ) +
  scale_y_log10(
    breaks = c(1e8, 1e9, 1e10, 1e11, 1e12),
    labels = c("100 Mio.", "1 Mrd.", "10 Mrd.", "100 Mrd.", "1 Bio.")
  ) +
  labs(
    title = "Marketcap vs. EBITDA nach Sektor (logarithmiert)",
    x = "EBITDA (in US-Dollar)",
    y = "Marketcap (in US-Dollar)"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_smooth(se = FALSE, color = "blue3", linewidth = 0.5)
```



Diese Grafik stellte die grösste Herausforderung dar, da sie drei stetige Variablen gleichzeitig visualisiert: Umsatzwachstum, Tageskurs und Marktkapitalisierung. Sie zeigt die Korrelation zwischen Umsatzwachstum und Tageskurs, wobei die Punkte nach Marktkapitalisierung gewichtet sind. Die Darstellung der Marktkapitalisierung war besonders schwierig, da die Werte stark streuen: Unternehmen unter 100 Milliarden US-Dollar wären ohne Anpassung kaum sichtbar gewesen. Um dies auszugleichen, wurde die Punktgrösse mit `scale_size_continuous()` angepasst, sodass auch kleinere Unternehmen in der Grafik erkennbar bleiben. Die x- und y-Achsen sind logarithmiert (`scale_x_log10()` und `scale_y_log10()`), um die grosse Spannweite der Werte angemessen darzustellen. Die Punktgrößen reichen von 1 bis 25, und die wichtigsten Werte werden mit Beschriftungen versehen, um die Lesbarkeit zu verbessern. Zusätzlich wurde eine Trendlinie (`geom_smooth()`) eingefügt, wobei die Standardabweichung entfernt wurde (`se = FALSE`). Die Grafik zeigt klar, wie Umsatzwachstum und Tageskurs zusammenhängen, und ermöglicht durch die Gewichtung nach Marktkapitalisierung einen differenzierten Blick auf Unternehmen unterschiedlicher Grösse.

```
a <- ggplot(data, mapping = aes(x = Currentprice, y = Revenuegrowth)) +
  geom_point(aes(size = Marketcap), alpha = 0.6) +
  scale_x_log10() +
  scale_y_log10(
    breaks = c(0.01, 0.10, 1.00),
    labels = c(1, 10, 100)) +
  scale_size_continuous(
    range = c(1, 25),
    breaks = c(1e11, 2e12, 3e12),
    labels = c("100 Mrd.", "2 Bio.", "3 Bio."),
    name = "Marktkapitalisierung (in US Dollar)"
  ) +
  theme_bw() +
  geom_smooth(aes(col = Marketcap), se = FALSE) +
  labs(title = "Umsatzwachstum vs. Tageskurs nach Marktkapitalisierung",
       x = "Tageskurs (in US-Dollar)",
       y = "Umsatzwachstum (in %)")

ggsave("mein_landscape_plot.pdf", plot = a, width = 11, height = 8)
```

Umsatzwachstum vs. Tageskurs nach Marktkapitalisierung

