

# Modellierung der Kaltmiete

Vergleich Frankfurt am Main und Leipzig

Henrik Popp, Kai Herbst, Manuel Zeh

2024-01-28

## Inhaltsverzeichnis

<b>Aufgabenstellung</b>	<b>1</b>
<b>Einleitung</b>	<b>2</b>
<b>Datenerhebung</b>	<b>2</b>
<b>Explorative Datenanalyse</b>	<b>3</b>
<b>Modellierung</b>	<b>15</b>
<b>Zusammenfassung</b>	<b>16</b>
<b>Quellen und Hilfsmittel</b>	<b>16</b>

## Aufgabenstellung

Abschnitt	Aufgabe	Reiner Textumfang	Erledigt
Einleitung	Auf inhaltliche Aufgabenstellung eingehen	0,5 - 1 Seiten	[ ]
Datenerhebung	Wie wurden die Daten erhoben? (Suchfilter, Sortierung)	1 - 3 Sätze	[ ]
Explorative Datenanalyse	Analyse + eventl. Datenvorverarbeitung	1 - 2 Seiten	[ ]
Modellierung	Modellierung + Interpretation	1 - 2 Seiten.	[ ]
Zusammenfassung	Gemeinsam kurz zentrale Ergebnisse zusammenfassen + Auf Grenzen der Analyse eingehen	0,5 - 1 Seiten.	[ ]

- Hier auch noch Literatur recherchieren:
  - <https://de.statista.com/statistik/daten/studie/258635/umfrage/bruttokaltmiete-bewohnter-wohnungen-in-deutschland-nach-bundeslaendern/>
  - [https://www.deutschlandatlas.bund.de/DE/Karten/Wie-wir-wohnen/040-Mieten.html#\\_6a54aw429](https://www.deutschlandatlas.bund.de/DE/Karten/Wie-wir-wohnen/040-Mieten.html#_6a54aw429)
  - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8053893/>
  - <https://de.statista.com/statistik/daten/studie/262508/umfrage/mietpreise-in-frankfurt-am-main/>
  - <https://de.statista.com/statistik/daten/studie/1312743/umfrage/mieten-in-leipzig-nach-dem-baualter-der-wohnung/>
  - <https://de.statista.com/statistik/daten/studie/535299/umfrage/mietpreise-auf-dem-wohnungsmarkt-in-leipzig/>
  - <https://de.statista.com/statistik/daten/studie/1312730/umfrage/entwicklung-der-angebotsmieten-in-leipzig/>
  - [https://link.springer.com/chapter/10.1007/978-3-658-11757-3\\_4](https://link.springer.com/chapter/10.1007/978-3-658-11757-3_4)
  - [https://www.ifo.de/DocDL/ifoDD\\_14-06\\_03-10.pdf](https://www.ifo.de/DocDL/ifoDD_14-06_03-10.pdf)

## Einleitung

In dieser Fallstudie sollen die Kaltmieten der beiden Städte Frankfurt am Main und Leipzig miteinander verglichen und modelliert werden. Ziel ist es, die verschiedenen möglichen Einflussfaktoren auf die Kaltmiete in den jeweiligen Städte zu bestimmen und eine Modellierung der Kaltmiete zu erstellen.

Zu Beginn wird auf die Datenerhebung eingegangen. Hier soll erklärt werden, woher die verarbeiteten Daten stammen und unter welchen Bedingungen die Daten erhoben wurden. Mit der explorativen Datenanalyse sollen dann die erhobenen Daten beschrieben und veranschaulicht werden. Hierbei wird die Vorverarbeitung der Daten beschrieben, im Anschluss wird mithilfe von Grafiken und dazugehörigen Interpretationen eine Datenanalyse erstellt. Dabei soll unter anderem herausgefunden werden, welche erhobenen Variablen den größten Einfluss auf die Kaltmiete einer Stadt haben oder wie hoch die eventuellen Unterschiede der Mieten in den beiden Städten sind. Den zentralen Teil des Dokuments stellt die Modellierung dar. Hier soll die Kaltmiete modelliert, also durch ein selbsterstelltes statistisches Modell dargestellt werden. Zudem wird das Modell interpretiert. Zum Abschluss werden die Ergebnisse in der Zusammenfassung aufgearbeitet und präsentiert.

## Datenerhebung

Die Datenerhebung fand ausschließlich über den Online-Marktplatz für Wohnungen und Häuser *ImmobilienScout24* statt. Die untersuchten Objekte wurden dabei auf den Immobilientyp *Wohnung* beschränkt, was als Suchkriterium in der Suchleiste des Portals eingestellt werden kann. Weitere Suchkriterien haben sich auf den *Ort*, in diesem Fall Frankfurt am Main und Leipzig, und auf den Objekttyp, hier *Mieten*, beschränkt. Weitere

Kriterien wie *Anzahl der Zimmer*, *Fläche* oder einem *maximalen Preis* wurden auf den Standardeinstellungen belassen. Anschließend wurden je Ort der Reihe nach bis zu 45 Objekte in der von ImmobilienScout24 generierten Reihenfolge überprüft und in eine Excel-Datei aufgenommen, die im Folgendem als Basis für die Auswertung dient.

Aufgenommen in die Datenbasis wurden dabei die folgenden Variablen: der *Ort*, die *Kaltniete* in Euro, die *Wohnfläche* in Quadratmetern, das Angebot eines *Parkplatz*, die *Etage*, Anzahl der *Zimmer*, Vorhandensein eines *Balkon*, das *Baujahr* des Objektes, sowie der entsprechende Link zur Anzeige und dessen Abrufdatum.

Für die nachfolgenden Auswertungen und Analysen lesen wir zunächst die Excel-Datei ein:

```
# Pfad zur Excel-Datei erstellen
pfad_mieten <- here("Mieten.xlsx")
# Daten einlesen
mieten <- read_excel(pfad_mieten)
```

Über die Ausgabe der ersten sechs Einträge erhalten wir einen Einblick in die Daten:

```
# Obere 6 Beobachtungen
head(mieten)
```

```
# A tibble: 6 x 12
  Ort    Kaltmiete Wohnflaeche Parkplatz Etage Zimmer Balkon Einbaukueche Heizung
<chr>   <dbl>      <dbl> <chr>   <chr> <dbl> <chr>   <chr>      <chr>
1 Fran~    1800        70   ja     1      2 ja     ja      Fußbod~
2 Fran~    1500        60   ja     1      1 ja     ja      Zentra~
3 Fran~    2650       146.   ja     1      3 ja     ja      Fußbod~
4 Fran~    1700        94   ja     2      3 nein   ja      Fußbod~
5 Fran~    2000       113.   ja     3      4 ja     ja      Fußbod~
6 Fran~    1700       84.8   ja     3      3 ja     ja      Fußbod~
# i 3 more variables: Baujahr <dbl>, Link <chr>, Abrufdatum <dtm>
```

## Explorative Datenanalyse

Zu Beginn der explorativen Datenanalyse muss geprüft werden, ob die in der Datenquelle enthaltenen Daten auf eine bestimmte Art und Weise vorverarbeitet oder angepasst werden müssen. Hierzu kann zunächst mit `str(mieten)` die Struktur des Datensatzes angezeigt werden.

```
str(mieten)
```

```
tibble [100 x 12] (S3: tbl_df/tbl/data.frame)
 $ Ort      : chr [1:100] "Frankfurt" "Frankfurt" "Frankfurt" "Frankfurt" ...
```

```

$ Kaltmiete      : num [1:100] 1800 1500 2650 1700 2000 1700 1480 2800 1080 2600 ...
$ Wohnflaeche    : num [1:100] 70 60 146 94 113 ...
$ Parkplatz      : chr [1:100] "ja" "ja" "ja" "ja" ...
$ Etage          : chr [1:100] "1" "1" "1" "2" ...
$ Zimmer         : num [1:100] 2 1 3 3 4 3 2 3 2 4 ...
$ Balkon         : chr [1:100] "ja" "ja" "ja" "nein" ...
$ Einbaukueche   : chr [1:100] "ja" "ja" "ja" "ja" ...
$ Heizung        : chr [1:100] "Fußbodenheizung" "Zentralheizung" "Fußbodenheizung" "Fußbodenheizung" ...
$ Baujahr        : num [1:100] 2022 1970 2017 NA 2015 ...
$ Link           : chr [1:100] "https://www.immobilienscout24.de/expose/136299839?referrer=R..."
$ Abrufdatum     : POSIXct[1:100], format: "2023-12-28" "2023-12-28" ...

```

```

miete_ffm <- subset(mieten, Ort == "Frankfurt")
miete_lpz <- subset(mieten, Ort == "Leipzig")

```

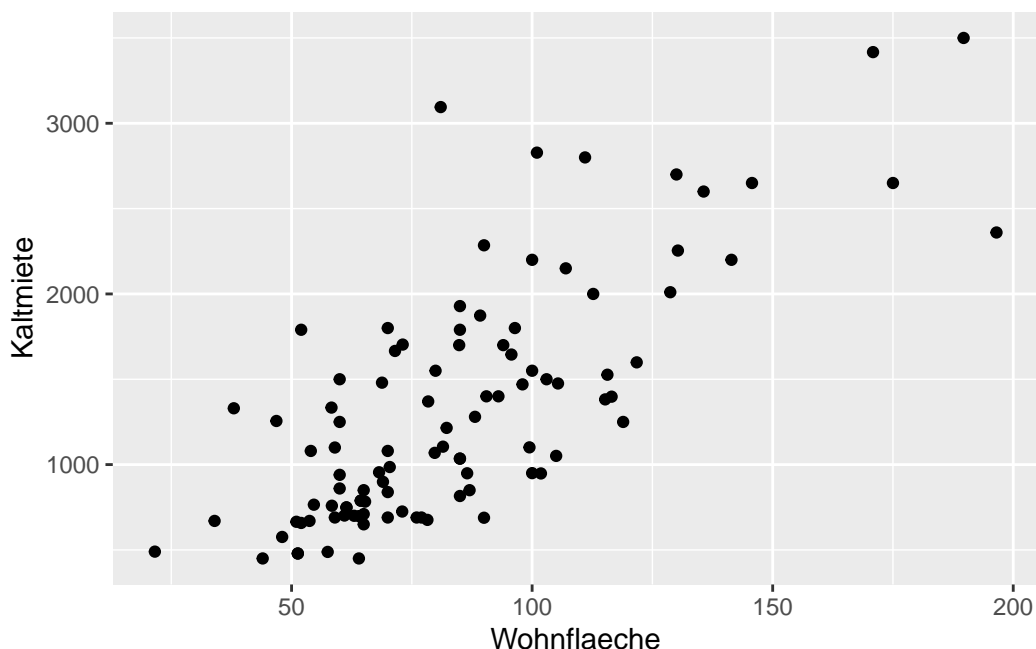
Es kann festgestellt werden, dass im Datensatz sowohl kategoriale nominale Variablen wie `Heizung` oder `Zimmer`, als auch metrische verhältnisskalierte Variablen wie `Kaltmiete` oder `Wohnflaeche` auftreten. Zunächst werden keine Variablen angepasst bzw. Werte ersetzt, da für die späteren Diagramme die kategorial nominalen Variablen als Achsenbeschriftung gut verwendet werden können.

Zuerst soll auf den Zusammenhang von `Kaltmiete` und `Wohnflaeche` geschaut werden, bei zwei metrisch verhältnisskalierten Variablen bietet sich dafür ein Scatterplot an.

```

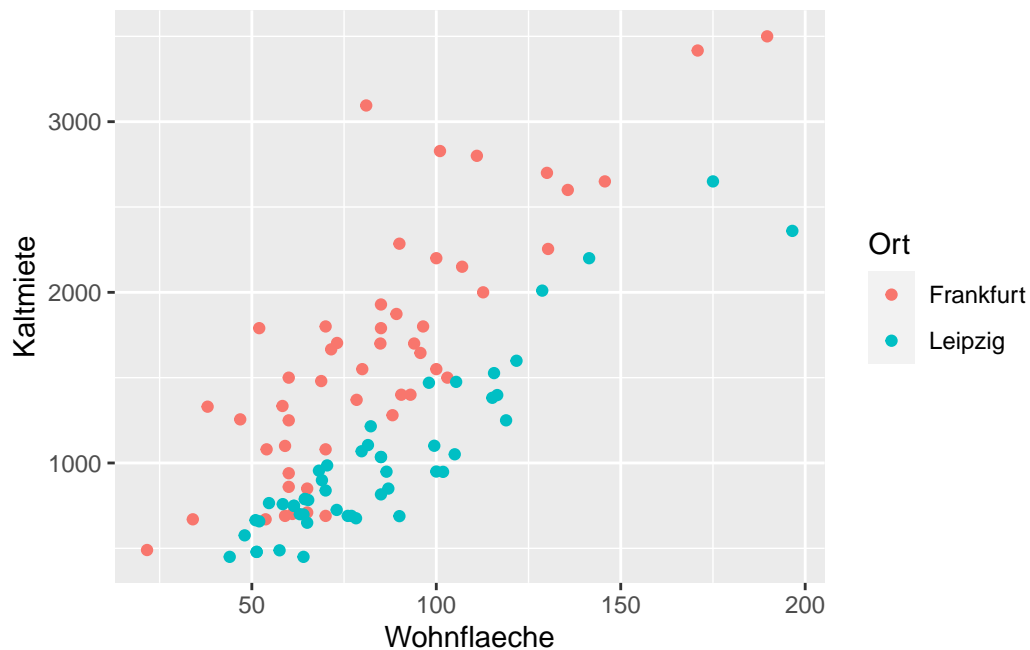
gf_point(Kaltmiete ~ Wohnflaeche, data = mieten)

```



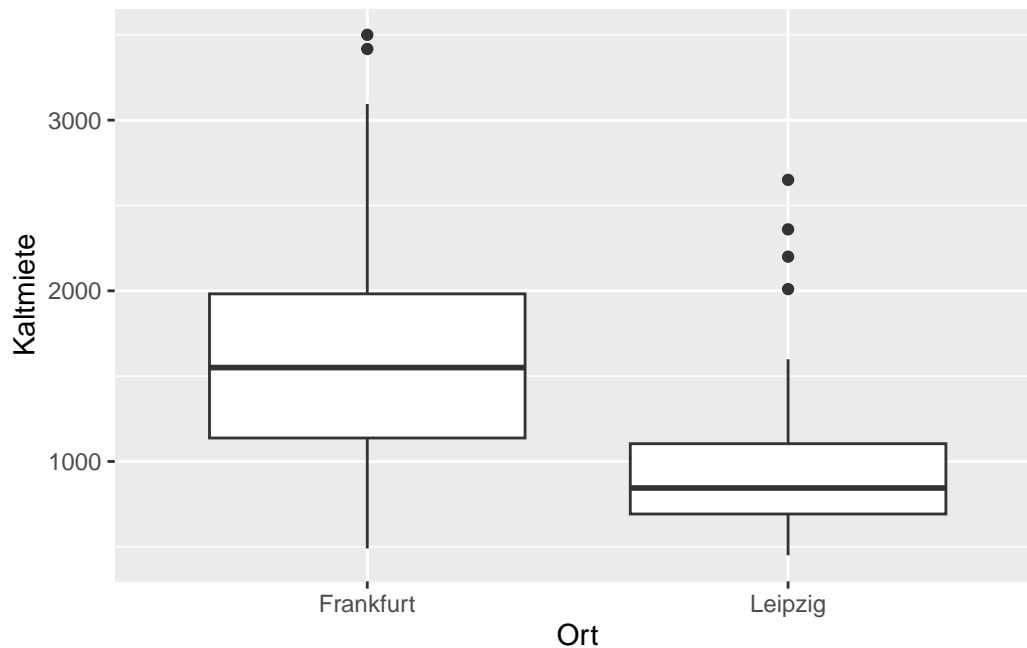
Grundsätzlich lässt sich ein positiver Zusammenhang zwischen **Kaltniete** und **Wohnflaeche** erkennen, wobei die Streuung der Kaltniete mit zunehmender Wohnfläche zunimmt. Nun muss dieses Diagramm jedoch um die Information des Ortes erweitert werden, um eine genauere Aussage treffen zu können. Dafür wird der Code um den Zusatz `color = ~ Ort` erweitert:

```
gf_point(Kaltniete ~ Wohnflaeche, data = mieten, color = ~ Ort)
```



Hier lässt sich nun erkennen, dass die erfassten Mieten im Datensatz in Frankfurt tendenziell höher sind, als in Leipzig. Bei vergleichbarer Wohnfläche liegen die gefärbten Punkte für Frankfurt stets über den Punkten von Leipzig. Um den Eindruck der Mietunterschiede zu festigen, kann ein Boxplot verwendet werden.

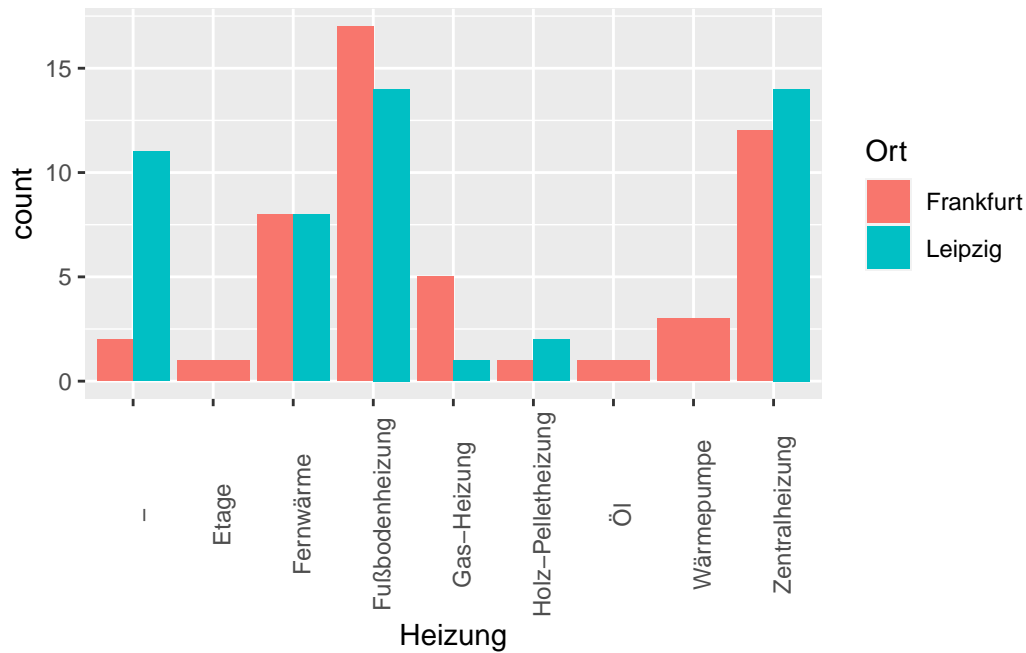
```
gf_boxplot(Kaltniete ~ Ort, data = mieten)
```



Das Boxplot zeigt, dass der Median für die Kaltmiete in Frankfurt deutlich über dem Median von Leipzig liegt. Zudem ist der 1,5-fache Interquartilsabstand bei Frankfurt größer als bei Leipzig, und die Whisker sind bei Frankfurt ebenfalls länger. Für Leipzig gibt es jedoch mehr Ausreißer, nämlich 4, verglichen mit den 2 Ausreißern von Frankfurt.

Es sollen nun auch die weiteren Variablen untersucht werden, angefangen mit der Variablen **Heizung**. Um die verschiedenen Ausprägungen zu vergleichen und ihre absolute Häufigkeit darzustellen, eignet sich ein Säulendiagramm:

```
# gf_boxplot(Kaltmiete ~ Heizung, data = mieten, color = ~ Ort)
# gf_point(Kaltmiete ~ Heizung, data = mieten, color = ~ Ort)
gf_bar( ~ Heizung, data = mieten, fill = ~ Ort, position = position_dodge()) + theme(ax
```

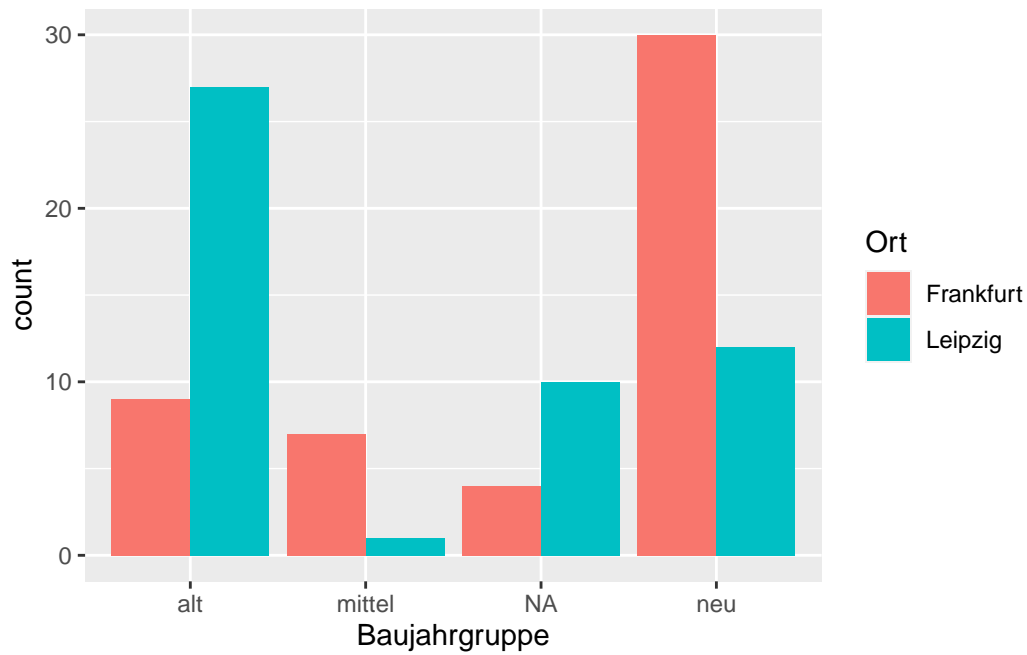


Das Säulendiagramm zeigt, dass die Fußbodenheizung in Frankfurt am weitesten verbreitet ist, gefolgt von der Zentralheizung und der Fernwärme. In Leipzig sind Gas-Heizung und Zentralheizung gleich häufig vertreten, gefolgt von der Fernwärme. In einigen Beobachtungen und häufiger in Leipzig als in Frankfurt, wurde der Heizungstyp nicht angegeben.

Bei der Variablen **Baujahr** handelt es sich hier um eine diskrete Variable. Aufgrund der Vielzahl an Jahren im Datensatz eignet sich jedoch die Verwendung des tatsächlichen Baujahres nicht, da die Darstellungen sonst sehr unübersichtlich werden. Stattdessen soll eine Klassifizierung in “alt - mittel - neu” vorgenommen werden, um die Baujahre zusammenzufassen.

```
mieten <- mieten %>%
  mutate(Baujahrgruppe = case_when(
    is.na(Baujahr) ~ "NA",
    as.integer(Baujahr) < 1970 ~ "alt",
    between(as.integer(Baujahr), 1970, 2000) ~ "mittel",
    TRUE ~ "neu"
  ))

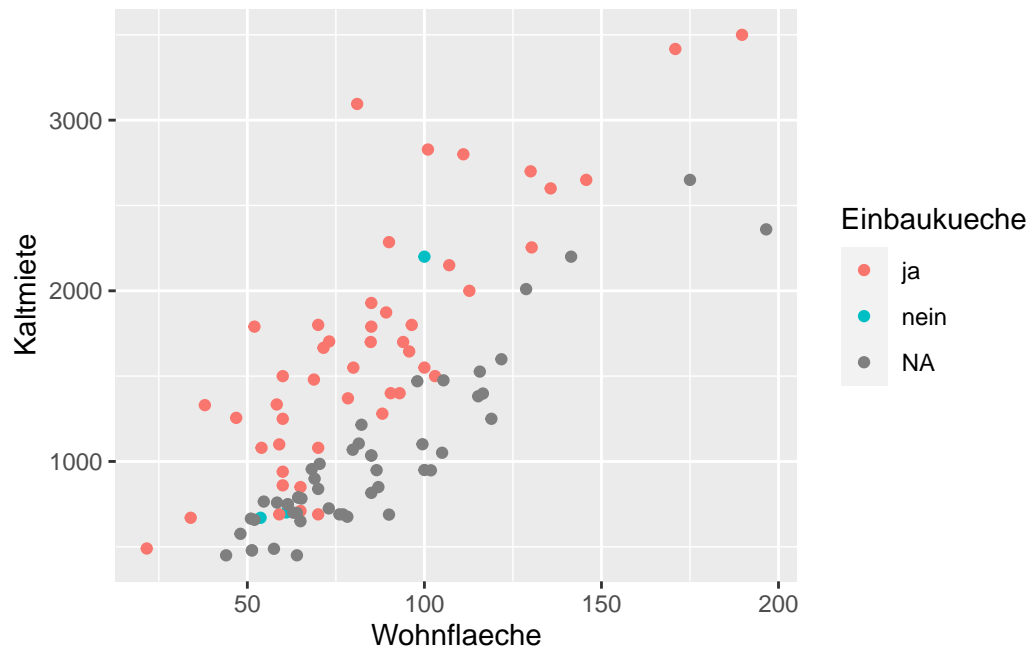
gf_bar( ~ Baujahrgruppe, data = mieten, fill = ~ Ort, position = position_dodge())
```



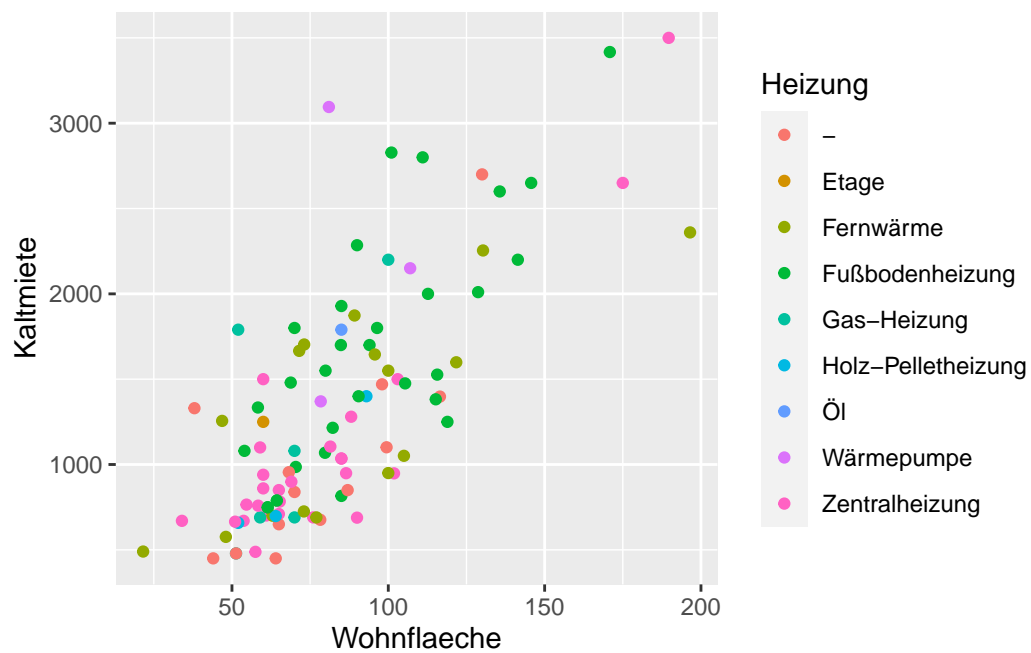
Im Säulendiagramm ist erkennbar, dass die meisten Beobachtungen in Frankfurt in die Kategorie “neu” (Baujahr > 2000) fallen, gefolgt von “alt” (Baujahr < 1970) und “mittel”. In Leipzig dominieren die Beobachtungen mit “alt”, dann kommen “neue” Baujahre. Es treten in beiden Städten nur wenige Beobachtungen mit Baujahren zwischen 1970 und 2000 auf, in Leipzig wurde das Baujahr häufiger nicht angegeben.

```
gf_point(Kaltmiete ~ Wohnflaeche, data = mieten, color = ~ Einbaukueche)
```

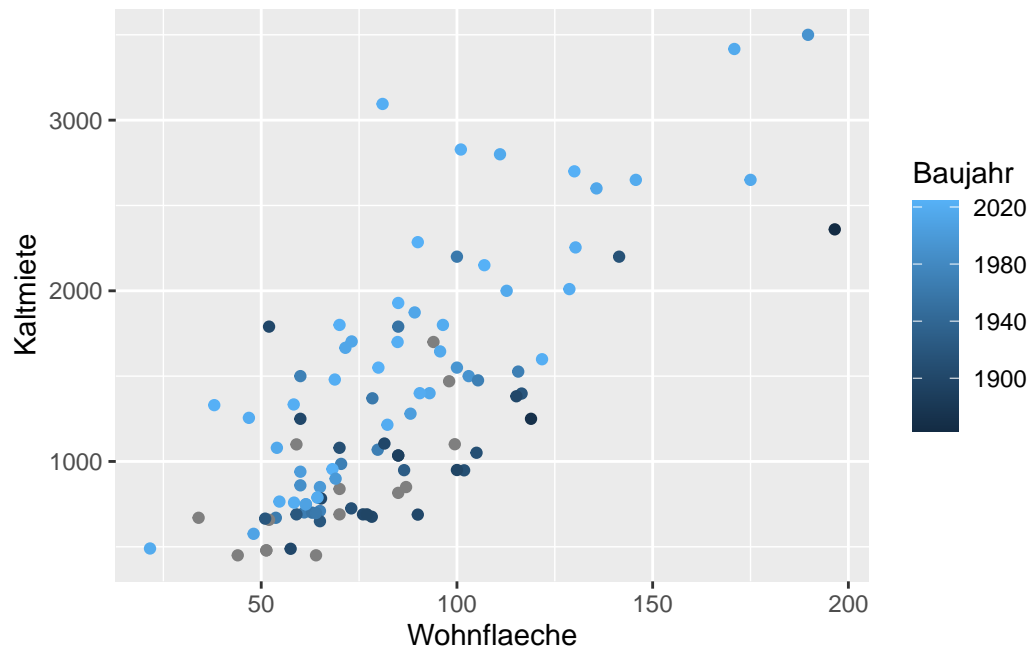




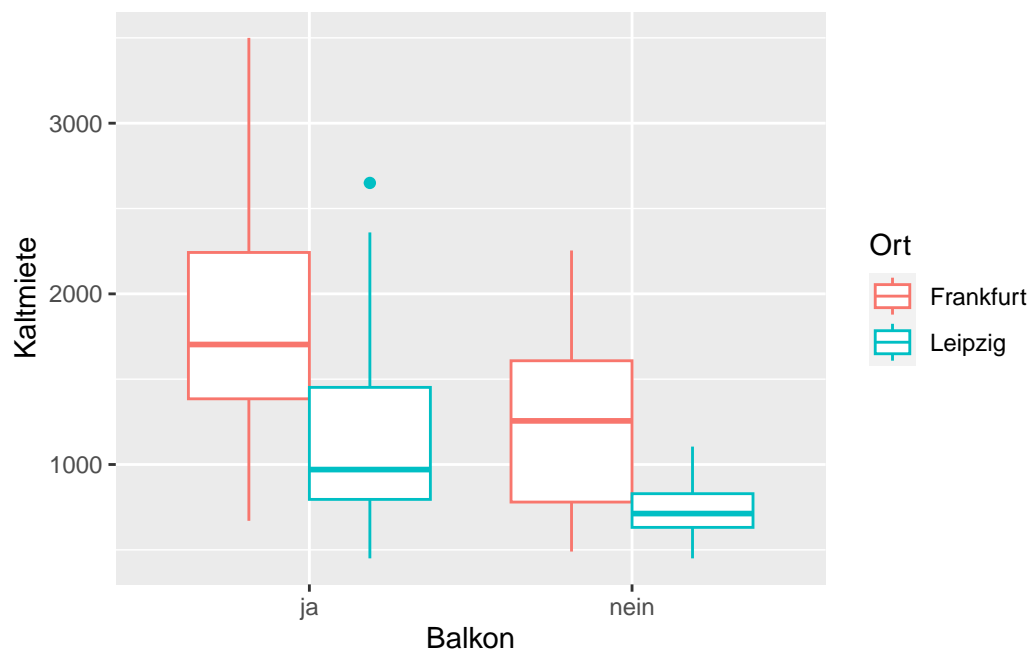
```
gf_point(Kaltmiete ~ Wohnflaeche, data = mieten, color = ~ Heizung)
```



```
gf_point(Kaltmiete ~ Wohnflaeche, data = mieten, color = ~ Baujahr)
```

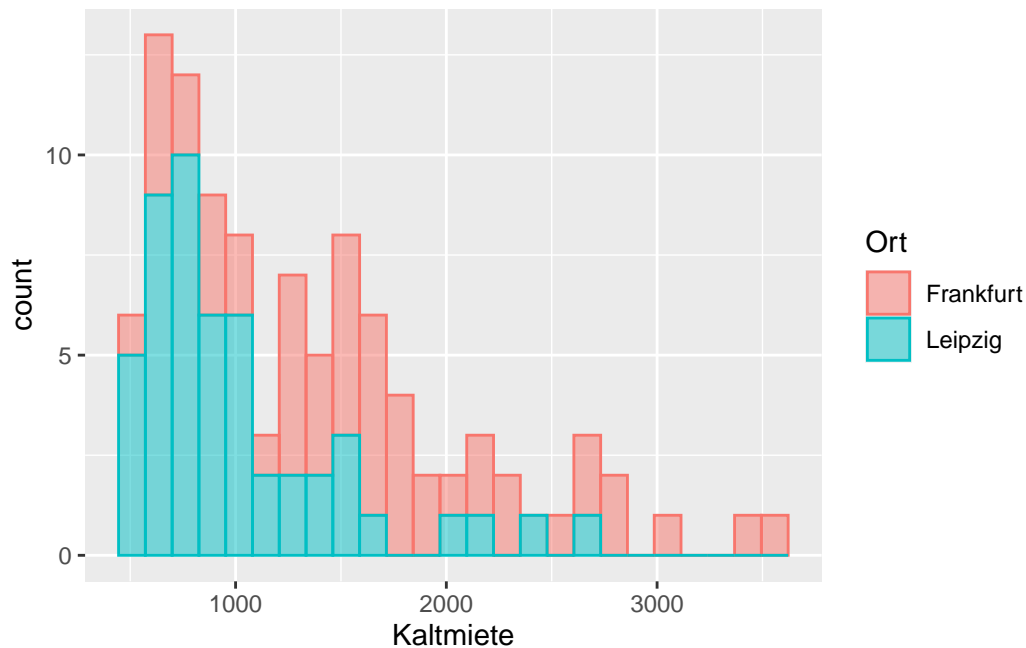


```
gf_boxplot(Kaltmiete ~ Balkon, data = mieten, color = ~ Ort)
```



```
gf_boxplot(Kaltmiete ~ Parkplatz, data = mieten, color = ~ Ort)
```





```
mean(~ Kaltmiete, data = miete_ffm)
```

```
[1] 1652.121
```

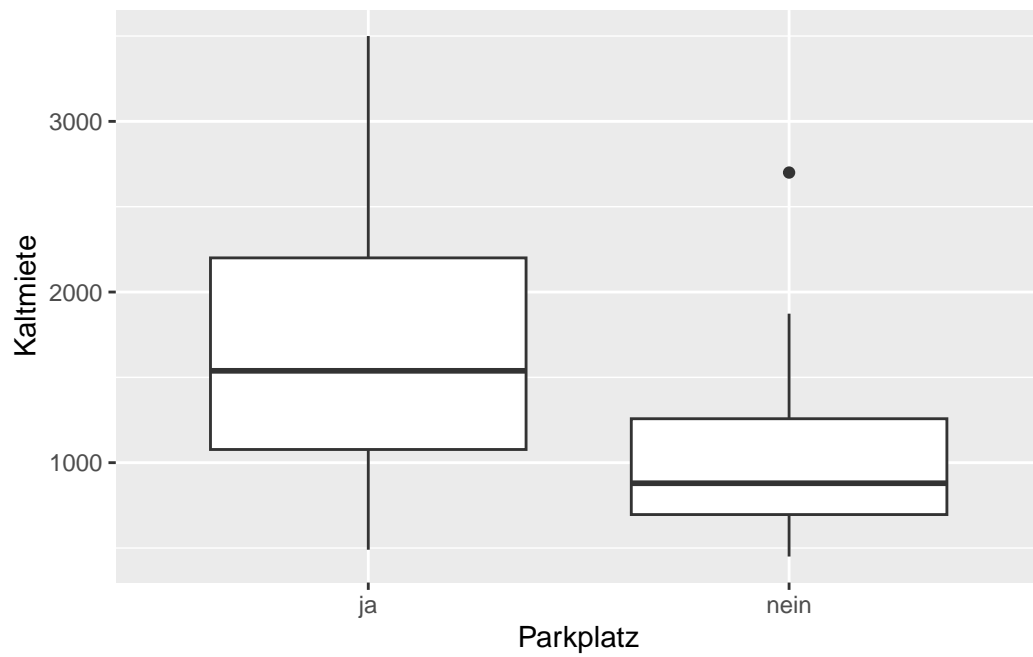
```
mean(~ Kaltmiete, data = miete_lpz)
```

```
[1] 1005.372
```

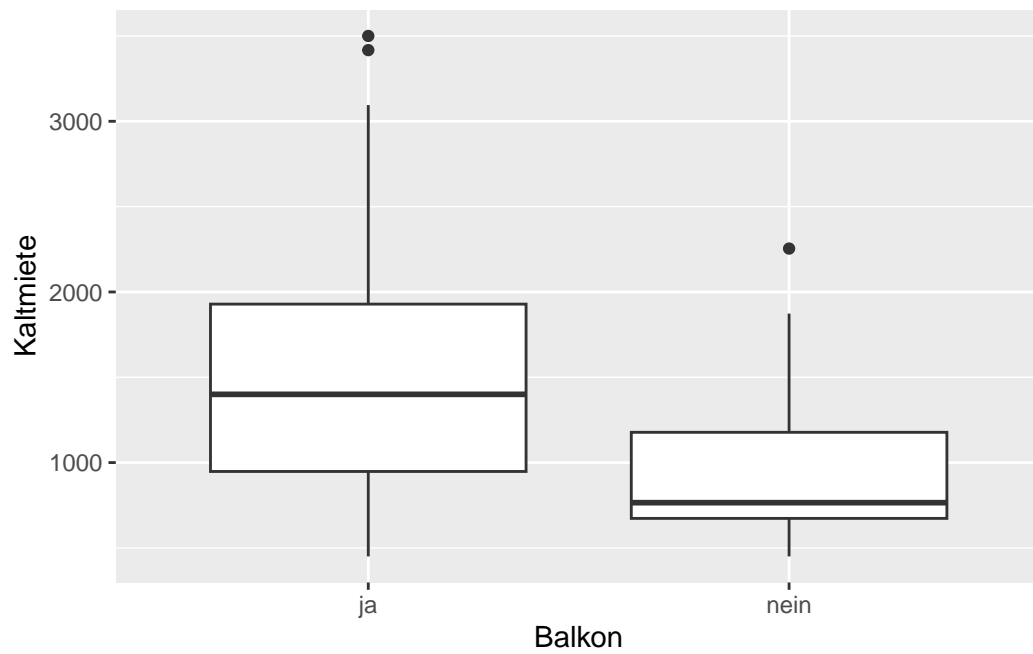
Weitere Ideen zur Analyse: - Kaltmiete pro Quadratmeter nach Stadt berechnen und als eigene Spalte pro Beobachtung hinzufügen + Stichpunkte zur Erklärung (Henrik) - Manuel: Kaltmiete zu Wohnfläche nach Stadt, von Henrik berechnet / Einbauküche / Heizung / Baujahr -> Einfluss auf Kaltmiete bzw. grafisch darstellen und analysieren -> Nach Stadt unterscheiden - Henrik: Kaltmiete bzw. Quadratmeterpreis nach Parkplatz / Etage / Zimmer / Balkon unterscheiden und analysieren - Eventuell die Baujahre in 3 Kategorien einteilen zur besseren Darstellung

- Wie unterscheiden sich die Städte, und wie verhält sich der Durchschnittspreis?
  - Boxplot mit Kaltmiete ~ Stadt
  - Boxplot mit Kaltmiete ~ Balkon oder Parkplatz mit Unterscheidung nach Städten
  - > Ggf. Grafik wieder rausnehmen wenn es keinen großen Unterschied gibt.
- Etage analysieren -> Einfluss auf Miete
- Wohnfläche analysieren
- Baujahr mit Histogramm darstellen -> Farbliche Unterteilung nach Städten

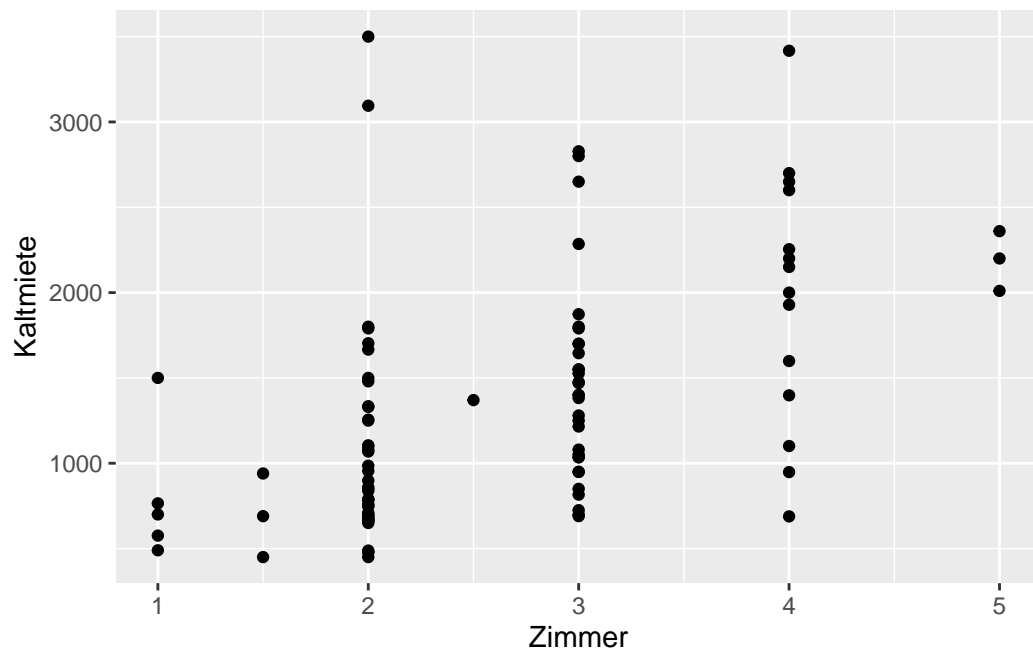
```
gf_boxplot(Kaltniete ~ Parkplatz, data = mieten)
```



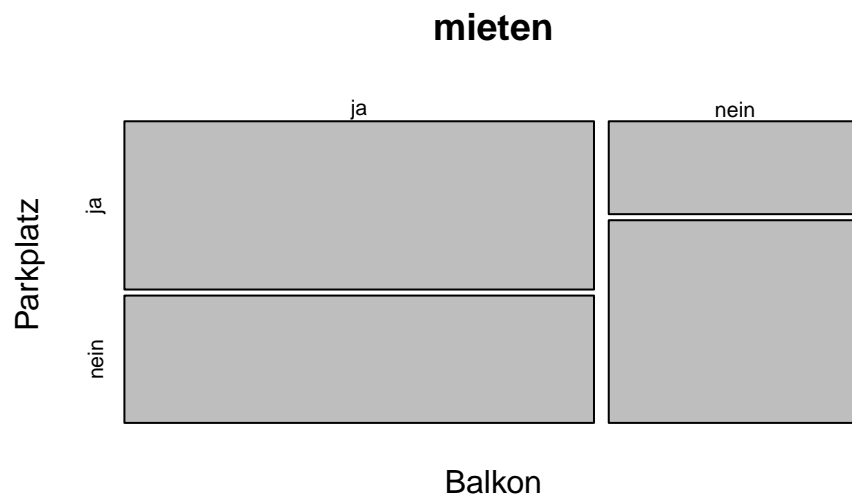
```
gf_boxplot(Kaltniete ~ Balkon, data = mieten)
```



```
gf_point(Kaltmiete ~ Zimmer, data = mieten)
```



```
mosaicplot(Balkon ~ Parkplatz, data = mieten)
```



Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## Modellierung

Modellieren Sie in diesem Abschnitt die Miete und interpretieren Sie Ihr Ergebnis.

Bei Einzelarbeiten sollte der reine Text (ohne Code, Abbildungen etc.) einen Umfang von ca. 0,5–1 Seiten haben, bei Gruppenarbeiten einen von ca. 1–2 Seiten.

Modell schätzen:

```
Modell_Mieten <- lm(Kaltemiete ~ 1, data = mieten)
summary(Modell_Mieten)
```

Call:

```
lm(formula = Kaltemiete ~ 1, data = mieten)
```

Residuals:

Min	1Q	Median	3Q	Max
-878.8	-572.3	-225.8	371.2	2171.2

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1328.75	70.11	18.95	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 701.1 on 99 degrees of freedom

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## **Zusammenfassung**

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

---

## **Quellen und Hilfsmittel**

Führen Sie hier die verwendeten Hilfsmittel sowie die verwendete Literatur auf.