



Technische
Universität
Braunschweig



Einführung in die Programmierung für Nicht-Informatiker

Hochschulstatistik

Lucy Thiele, Melina Sablotny, Julia Limbach, Marleen Adolphi, Nele Grundke | 18.02.2025

Visualisierung Hochschulstatistik

1. Geeignete grafische Darstellung der zeitlichen Entwicklung der Studierendenzahlen und Frauenquote in den verschiedenen Studiengängen
2. Geeignete grafische Darstellung der zeitlichen Entwicklung der absoluten Zahlen und der Frauenquote des wissenschaftlichen Hochschulpersonals für die verschiedenen Fachbereiche
3. Geeignete grafische Darstellung der zeitlichen Entwicklung der Hochschulausgaben der einzelnen Bundesländer (sowohl die Gesamtausgaben als auch die Ausgaben der einzelnen Fachbereiche sowie deren Anteil an den Gesamtausgaben)
4. Bonus: Gegenüberstellung der Entwicklungen von Studierendenzahlen, Personalzahlen und Ausgaben

Aufgabe 1: Zeitliche Entwicklung der Studierenden und der Frauenquote



Wie haben sich die absoluten Zahlen der Studierenden und die Frauenquote in den Fachbereichen Betriebswirtschaftslehre (BWL)/ Wirtschaftswissenschaften und Elektrotechnik in den Jahren 1998 bis 2021 geändert?

Aufgabe 1: Annahmen

Es gibt mehr Studierende in dem Studiengang Betriebswirtschaftslehre als in dem Studiengang Elektrotechnik.

"Frauen [entscheiden sich] nach wie vor seltener für ein Studium in einem MINT-Fach als Männer." (Statistisches Bundesamt, 2018).

Die Frauenquote in BWL deutlich höher ist, als die Frauenquote in Elektrotechnik.

Der Frauenanteil in MINT-Fächern steigt (Statistisches Bundesamt, 2024).

Die Frauenquote steigt in Elektrotechnik stärker als im Studiengang BWL.

Aufgabe 1: Code

```
# Berechnung der Frauenquote
df['Frauenquote'] = (df['Weiblich'] / df['Insgesamt']) * 100

# Berechnung prozentuale Veränderung
def berechne_prozentuale_veraenderung(df, studiengang):
    frauenquote = df[df['Studiengang'] == studiengang].set_index('Jahr')['Frauenquote']
    startwert = frauenquote.iloc[0]
    endwert = frauenquote.iloc[-1]
    absolute_veraenderung = endwert - startwert
    relative_veraenderung = ((endwert - startwert) / startwert) * 100

    return absolute_veraenderung, relative_veraenderung, endwert

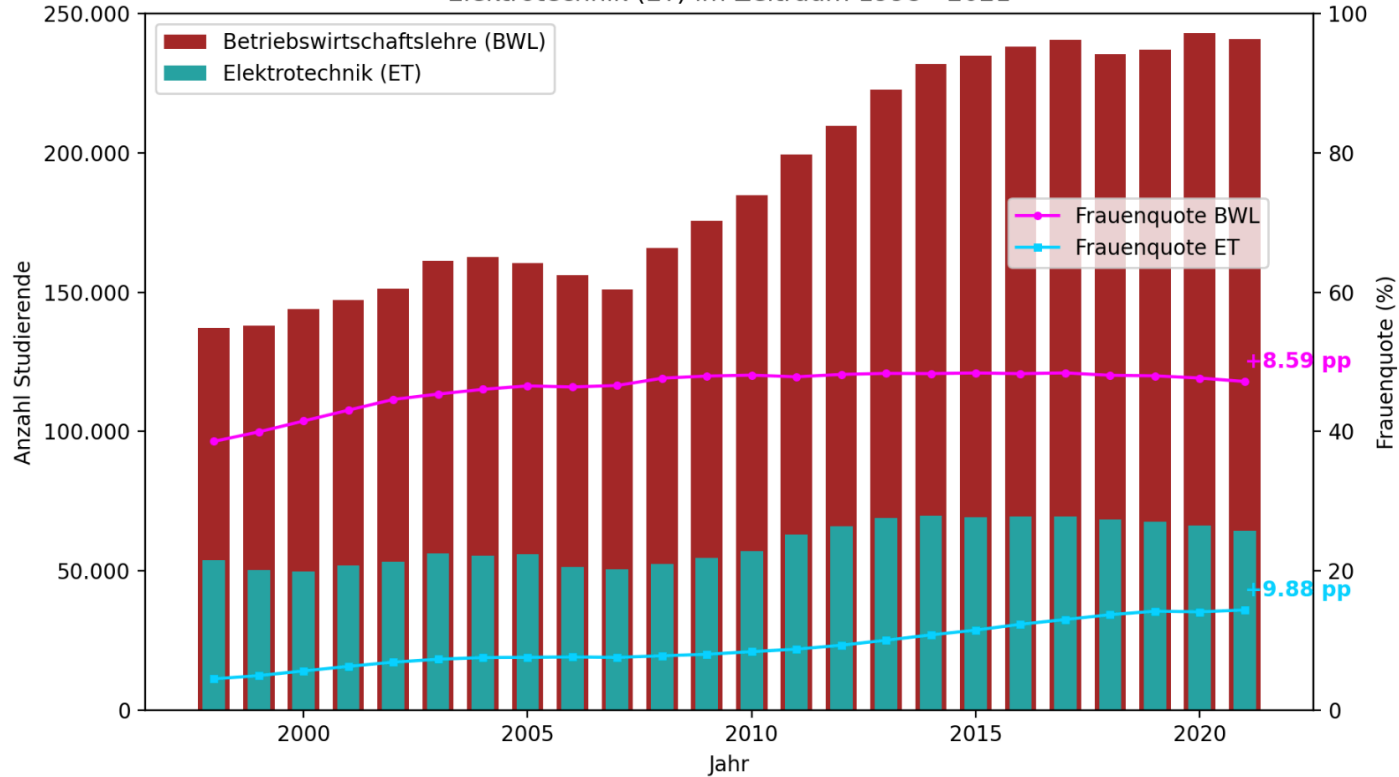
# Berechnung der prozentualen Veränderung für beide Studiengänge
abs_veraenderung_bwl, rel_veraenderung_bwl, endwert_bwl = berechne_prozentuale_veraenderung(df, 'Betriebswirtschaftslehre')
abs_veraenderung_etch, rel_veraenderung_etch, endwert_etch = berechne_prozentuale_veraenderung(df, 'Elektrotechnik')

# Berechnung Frauen pro zusätzlichem Studierenden pro Fach
bwl_gesamt = df[df['Studiengang'] == 'Betriebswirtschaftslehre'].groupby('Jahr')['Insgesamt'].sum()
etch_gesamt = df[df['Studiengang'] == 'Elektrotechnik'].groupby('Jahr')['Insgesamt'].sum()

bwl_frauen = df[df['Studiengang'] == 'Betriebswirtschaftslehre'].groupby('Jahr')['Weiblich'].sum()
etch_frauen = df[df['Studiengang'] == 'Elektrotechnik'].groupby('Jahr')['Weiblich'].sum()

frauen_pro_studierenden_bwl = bwl_frauen.diff().sum() / bwl_gesamt.diff().sum()
frauen_pro_studierenden_etch = etch_frauen.diff().sum() / etch_gesamt.diff().sum()
```

Studierendenzahlen und Frauenquote in den Studiengängen Betriebswirtschaftslehre (BWL) und Elektrotechnik (ET) im Zeitraum 1998 - 2021



Frauen pro
zusätzlichem
Studierenden in
BWL: **0.59**

Frauen pro
zusätzlichem
Studierenden in
Elektrotechnik:
0.65

Aufgabe 2: Zeitliche Entwicklung des Personals und der Frauenquote



Wie haben sich die absoluten Zahlen des Personals und die Frauenquote in den Fachbereichen Wirtschaftswissenschaften und Elektrotechnik in den Jahren 1998 bis 2020 geändert?

Aufgabe 2: Annahmen

- Es gibt mehr Personal im Bereich Wirtschaftswissenschaften, als im Bereich Elektrotechnik.
- Die Frauenquote in den Wirtschaftswissenschaften ist deutlich höher, als die Frauenquote in Elektrotechnik.
- Die Frauenquote steigt in den Wirtschaftswissenschaften stärker als bei Elektrotechnik.
- Im Jahr 2008 startete das Projekt *Komm, mach MINT*, welches vom Bundesministerium von Bildung und Forschung gefördert wurde. Daraufhin ist ein Anstieg des weiblichen Personals zu erkennen.

Aufgabe 2: Code

```
# Filter anlegen
insgesamt = personal['Geschlecht'] == 'insgesamt'
weiblich = personal['Geschlecht'] == 'w'
jahre = personal.columns[3:].astype(int) # Jahre als Ganzzahlen konvertieren

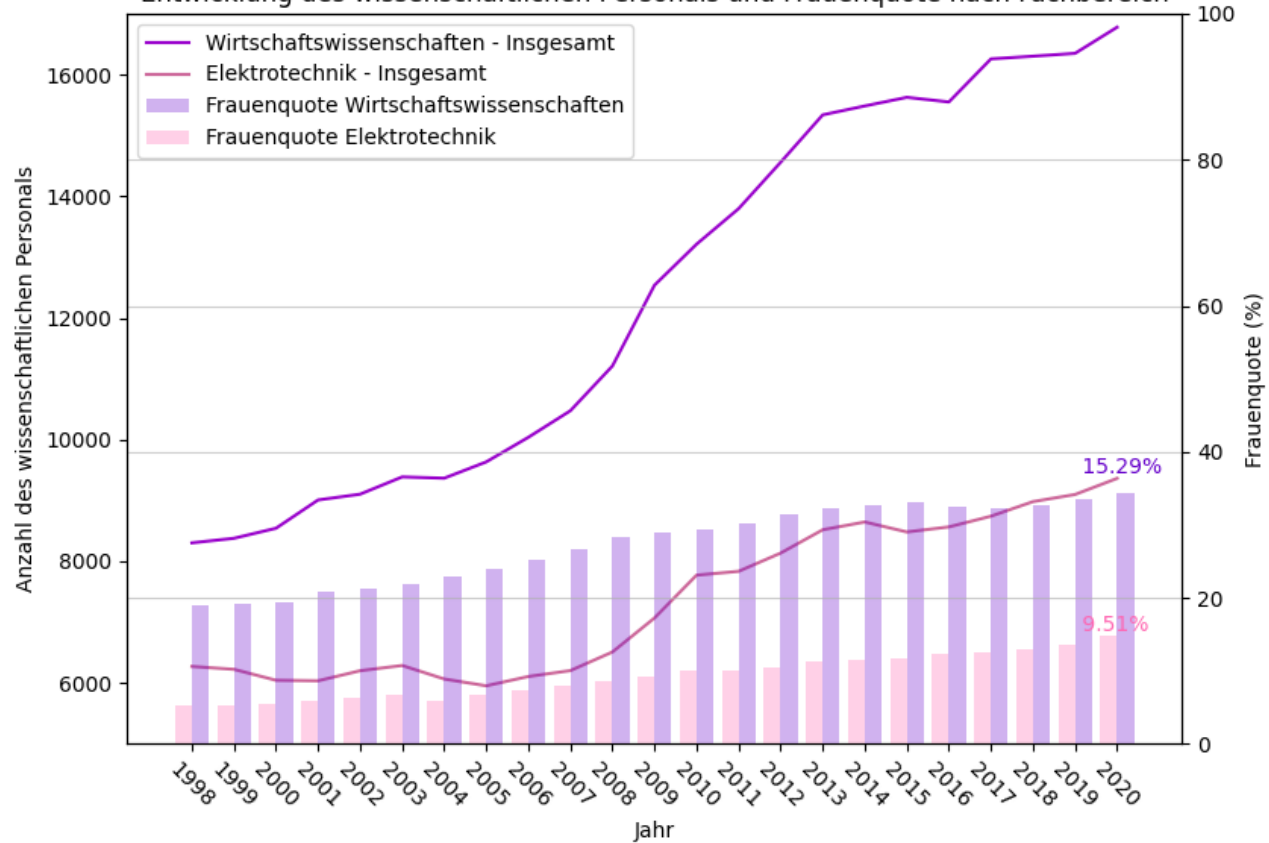
elektro = personal['Fachbereich'] == 'Elektro- und Informationstechnik'
frauenquote_elektro = (personal[weiblich & elektro][jahre.astype(str)].values /
                      personal[insgesamt & elektro][jahre.astype(str)].values) * 100

wirtschaft = personal['Fachbereich'] == 'Wirtschaftswissenschaften'
frauenquote_wirtschaft = (personal[weiblich & wirtschaft][jahre.astype(str)].values /
                          personal[insgesamt & wirtschaft][jahre.astype(str)].values) * 100

frauenquote = pd.DataFrame({
    "Jahr": jahre,
    "Elektrotechnik": frauenquote_elektro.flatten(),
    "Wirtschaftswissenschaften": frauenquote_wirtschaft.flatten()
})

# Berechnung der prozentualen Veränderung
änderung_elektro = frauenquote["Elektrotechnik"].iloc[-1] - frauenquote["Elektrotechnik"].iloc[0]
änderung_wirtschaft = frauenquote["Wirtschaftswissenschaften"].iloc[-1] - frauenquote["Wirtschaftswissenschaften"].i
```

Entwicklung des wissenschaftlichen Personals und Frauenquote nach Fachbereich



Lucy Thiele, Melina Sablotny, Julia Limbach, Marleen Adolphi, Nele Grundke | 18.02.2025

Aufgabe 3: Zeitliche Entwicklung der Hochschulausgaben



Wie haben sich die Hochschulausgaben in den einzelnen Bundesländern von 2006 bis 2020 entwickelt, und welche Faktoren beeinflussen diese Entwicklung?

Aufgabe 3: Annahmen

Hochschulausgaben steigen mit der Zeit

Die Anzahl der Hochschulen beeinflusst die Höhe der Ausgaben

Die Bundesländer im Westen und Süden haben tendenziell höhere Ausgaben als die im Norden und Osten

Im Süden und Osten liegen die Ausgaben für Ingenieurwissenschaften tendenziell höher als für Rechts,-Wirtschafts- und Sozialwissenschaften

Die Verteilung der Anteile der Fachbereiche der Bundesländer (innerhalb ihrer Region) an den Gesamtausgaben ist nicht immer an der Höhe dieser zu messen

Aufgabe 3: Code - Vergleich der Bundesländer

```
# Daten einlesen
data = pd.read_csv('ausgaben-hochschulen-21371-0002.csv')

# Daten filtern
data_filtered = data[data["Fachbereich"] == "Insgesamt"]
data_filtered = data_filtered.set_index("Jahr")
data_filtered = data_filtered.drop(columns=["Fachbereich"])
data_filtered = data_filtered / 1e6 # Umrechnung in Millionen €

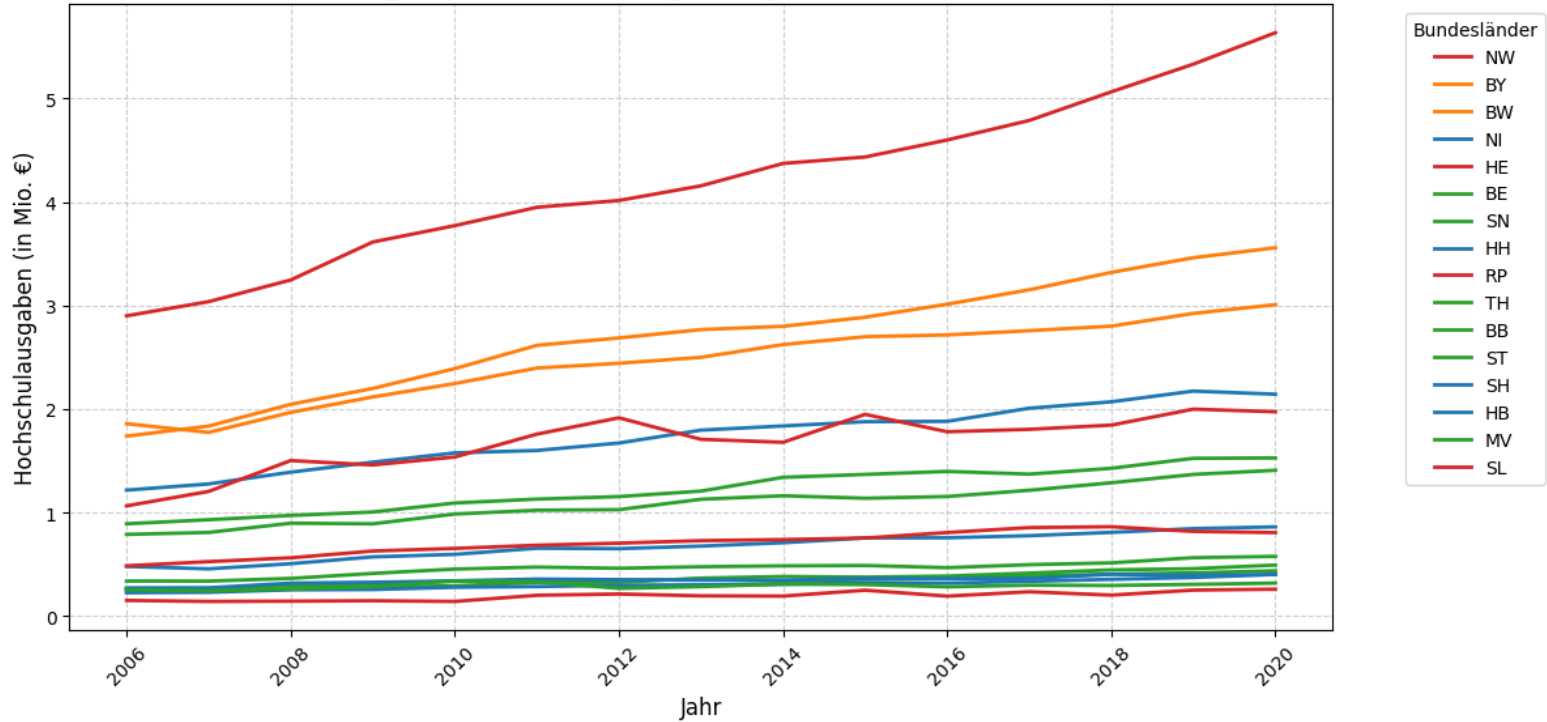
# Farben nach Regionen definieren
farben = {
    "SH": "#1f77b4", "HH": "#1f77b4", "HB": "#1f77b4", "NI": "#1f77b4", # Blau (Nord)
    "BY": "#ff7f0e", "BW": "#ff7f0e", # Orange (Süd)
    "MV": "#2ca02c", "BE": "#2ca02c", "BB": "#2ca02c", "ST": "#2ca02c", "SN": "#2ca02c",
    "NW": "#d62728", "HE": "#d62728", "RP": "#d62728", "SL": "#d62728" # Rot (West)
}

# Sortierung der Bundesländer nach ihrem Endwert im letzten Jahr
sorted_columns = data_filtered.iloc[-1].sort_values(ascending=False).index

# Berechnung der prozentualen Veränderung
percentage_change = ((data_filtered.loc[2020] - data_filtered.loc[2006]) / data_filtered.loc[2006]) * 100
```

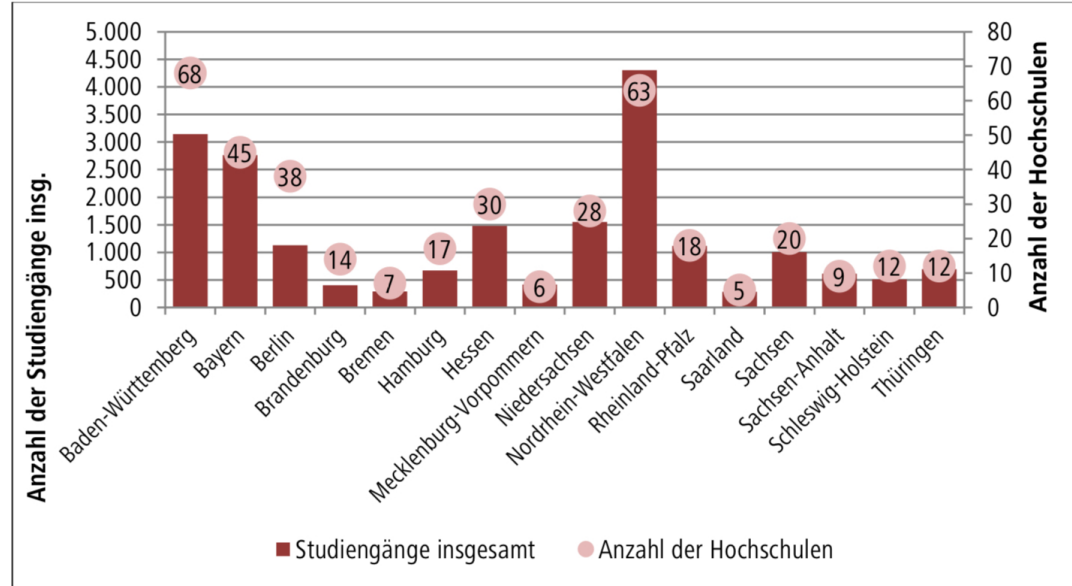
Ergebnisse Bundeslandvergleich

Entwicklung der Hochschulausgaben pro Bundesland (2006-2020)



Hochschulverteilung

Diagramm 1.5 Studienangebote nach Bundesländern im Wintersemester 2020/2021



Quelle: HRK-Hochschulkompass, 1.9.2020 (WiSe 2020/2021)

Aufgabe 3: Code - Vergleich der Fachbereiche in Regionen

```
# Auswahl der relevanten Fachbereiche
df_filtered = df[df["Fachbereich"].isin(["Wirtschaftswissenschaften", "Elektrotechnik"])]

# Aggregation der Daten nach Regionen
df_regions = df_filtered.copy()
for region, states in regions.items():
    df_regions[region] = df_regions[states].sum(axis=1)

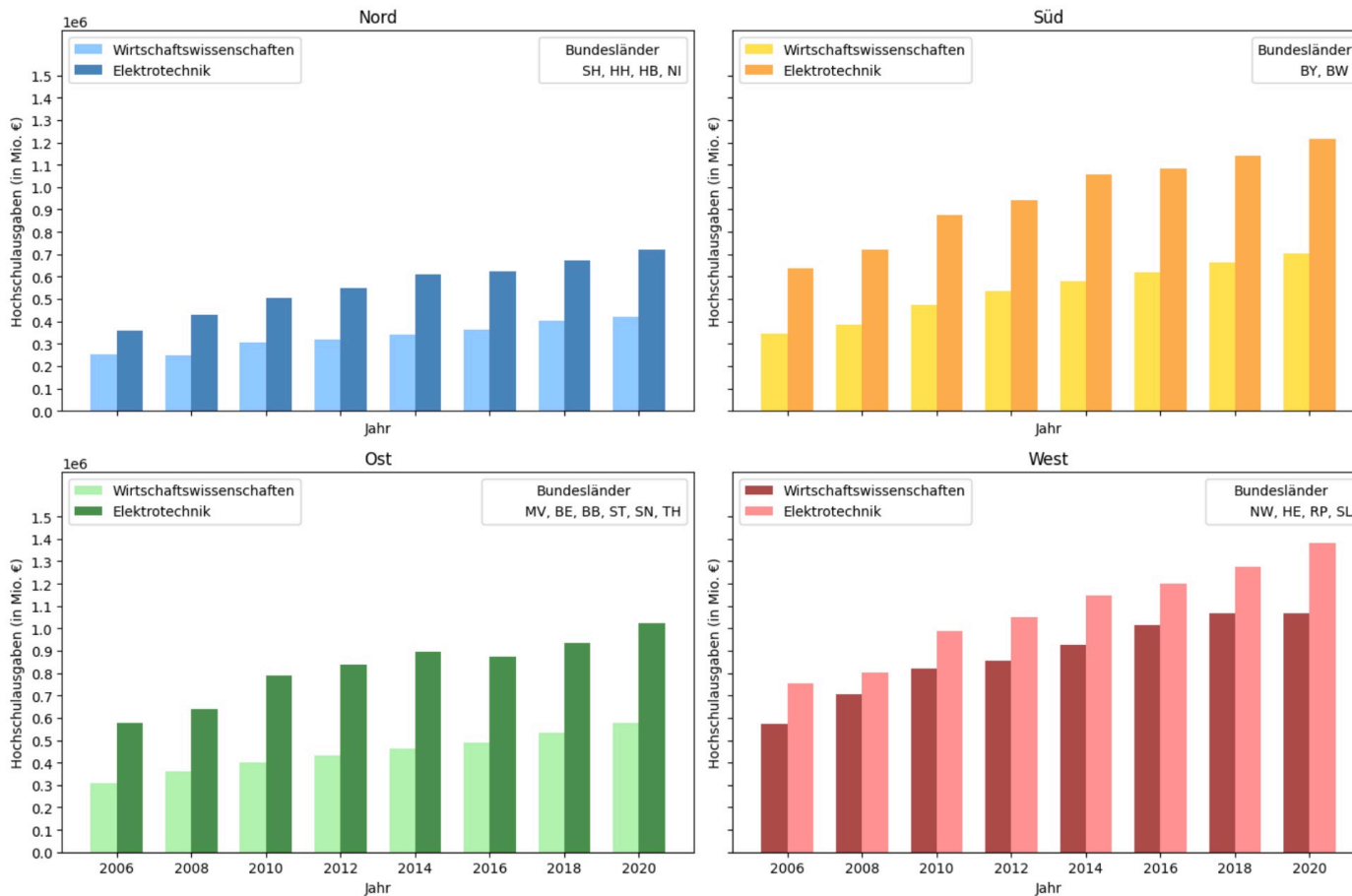
df_regions = df_regions.melt(id_vars=["Jahr", "Fachbereich"], value_vars=regions.keys(),
                             var_name="Region", value_name="Ausgaben")

# Plotten der Daten
fig, axes = plt.subplots(2, 2, figsize=(14, 10), sharex=True, sharey=True)
fig.suptitle("Entwicklung der Hochschulausgaben in den Fachbereichen (2006-2020)", y=0.95)

bar_width = 0.7 # Breite der Balken

# Regionen in Subplots aufteilen
region_list = list(regions.keys())
for i, ax in enumerate(axes.flat):
    region_name = region_list[i]
    region_data = df_regions[df_regions["Region"] == region_name]
```


Entwicklung der Hochschulausgaben in den Fachbereichen (2006-2020)



Lucy Thiele, Melina Sablotny, Julia Limbach, Marleen Adolphi, Nele Grundke | 18.02.2025

Aufgabe 3 (Anteil Fachbereiche): Code für das Beispiel Süden

```
#Anteile berechnen (vorher sueden_bundeslaender definiert als 'BY' und 'BW')
for bundesland in sueden_bundeslaender:
    prozentuale_anteile = []

    for jahr in df_filtered['Jahr'].unique():
        df_jahr = df_filtered[df_filtered['Jahr'] == jahr]
        gesamt_ausgaben = df_jahr[bundesland].sum() # Gesamtausgaben im aktuellen Jahr für
        anteile = (df_jahr[bundesland] / gesamt_ausgaben) * 100 #Prozentualer Anteil jeder
        prozentuale_anteile.append(anteile.values)

#Umwandlung in DataFrame
fachbereiche = df_filtered['Fachbereich'].unique()
prozentuale_anteile_df = pd.DataFrame(prozentuale_anteile, index=df_filtered['Jahr']).

#Farben Fachbereiche
fachbereich_colors = generate_colors(len(fachbereiche))

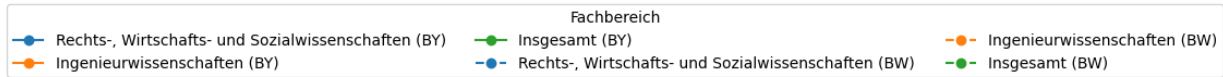
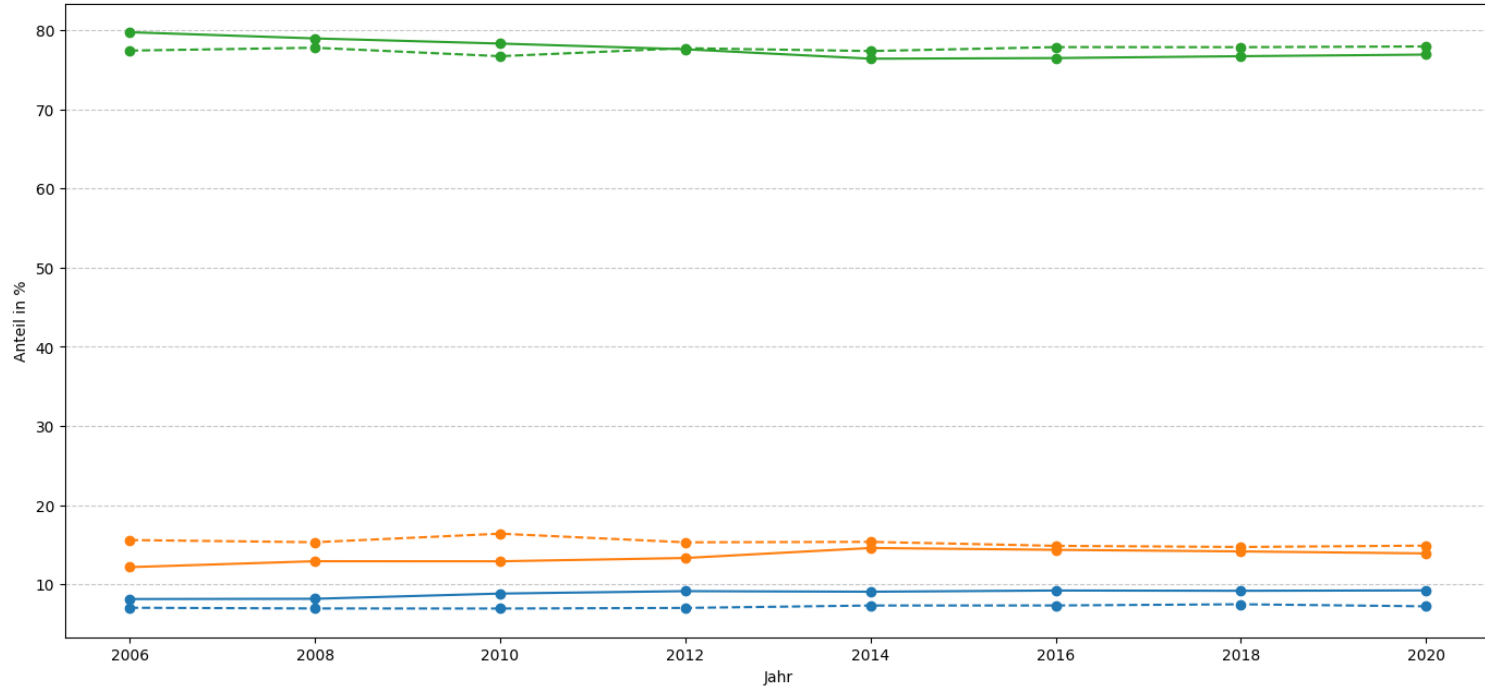
#Plot für jeden Fachbereich im Bundesland
for idx, fachbereich in enumerate(fachbereiche):
    farbe = fachbereich_colors[idx]
    linienstil = linienstile[bundesland]

    linie, = plt.plot(prozentuale_anteile_df.index, prozentuale_anteile_df[fachbereich],
                     linestyle=linienstil, marker='o', label=f'{fachbereich} ({bundesland})')

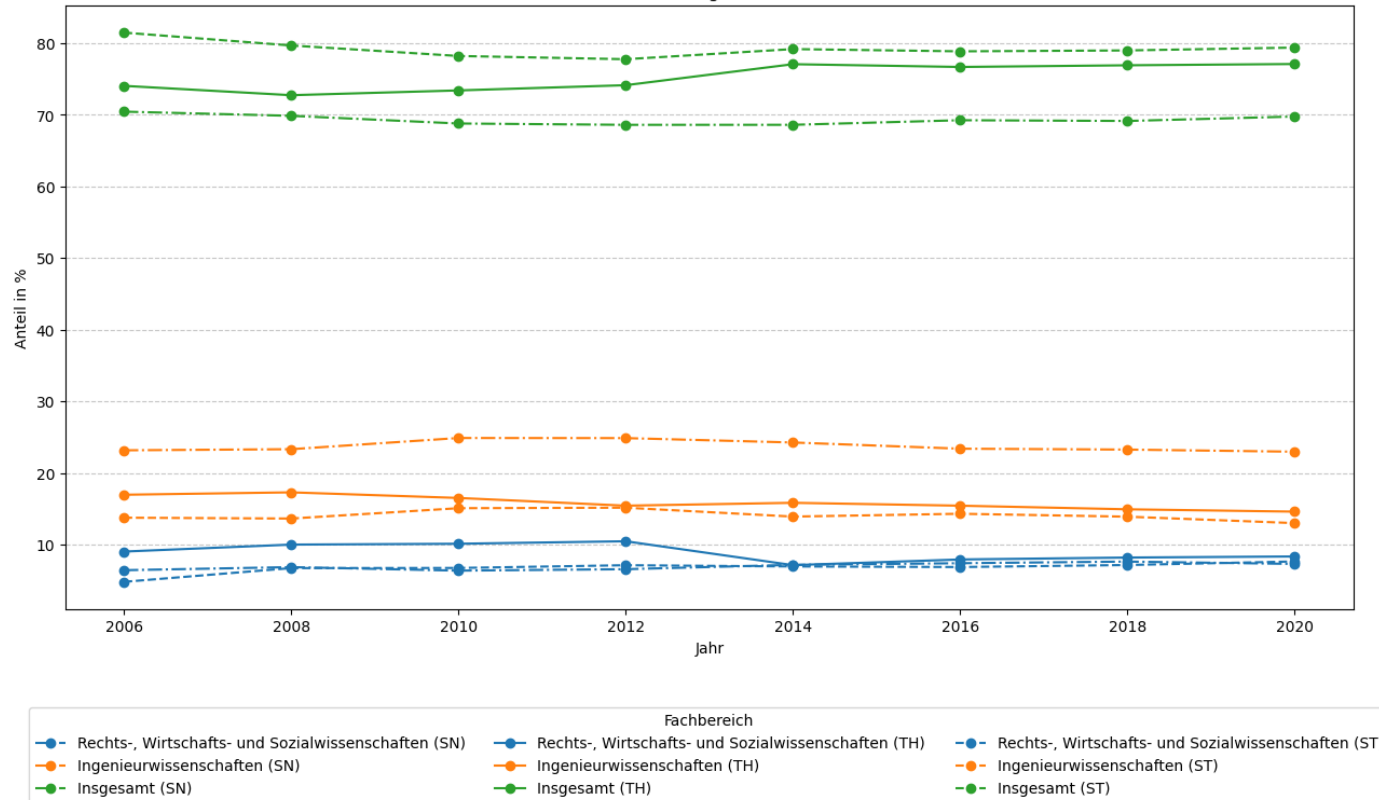
    linien.append(linie)
    labels.append(f'{fachbereich} ({bundesland})')

plt.title('Anteile der Fachbereiche an den Gesamtausgaben im Süden (BY, BW) von 2006-2020')
plt.xlabel('Jahr')
plt.ylabel('Anteil in %')
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

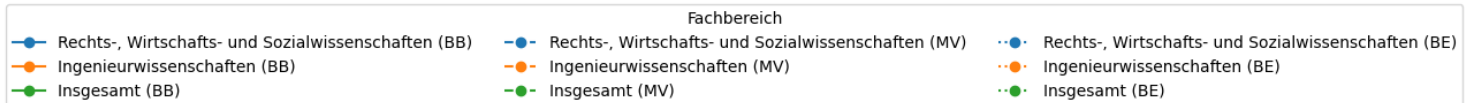
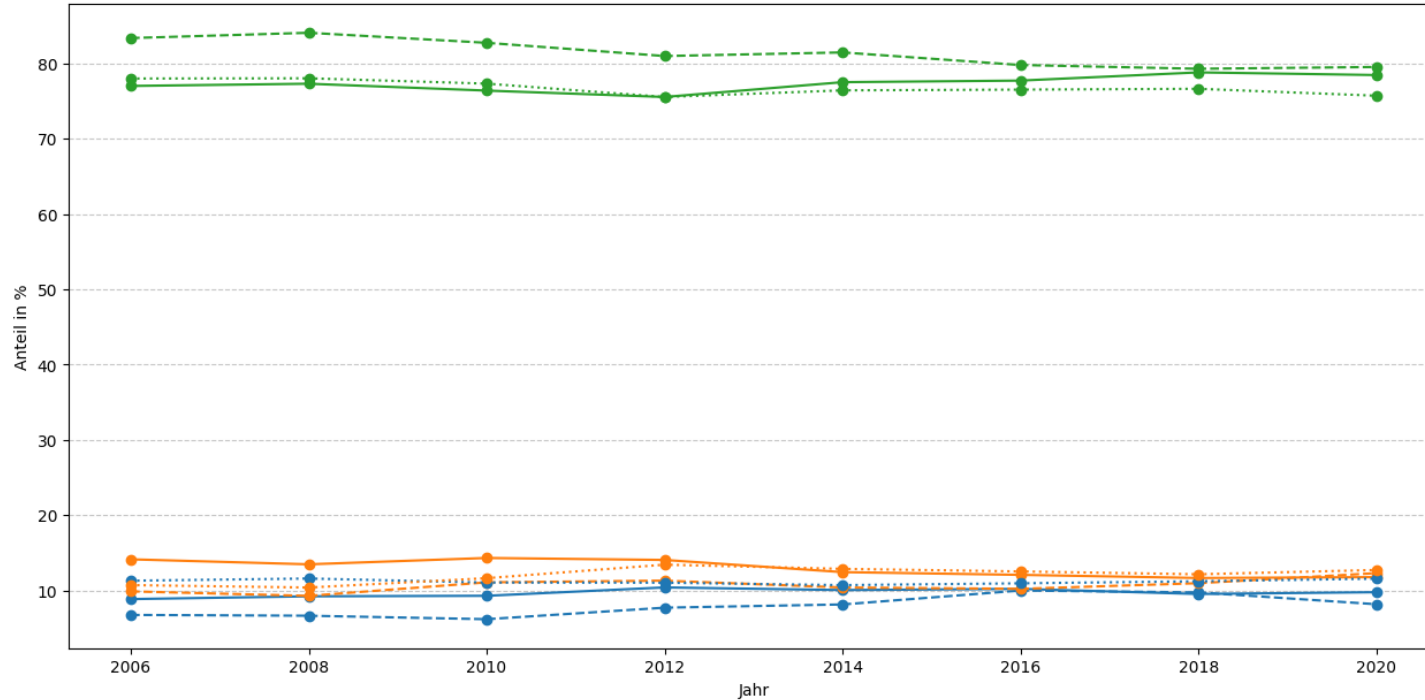
Anteile der Fachbereiche an den Gesamtausgaben im Süden (BY, BW) von 2006-2020



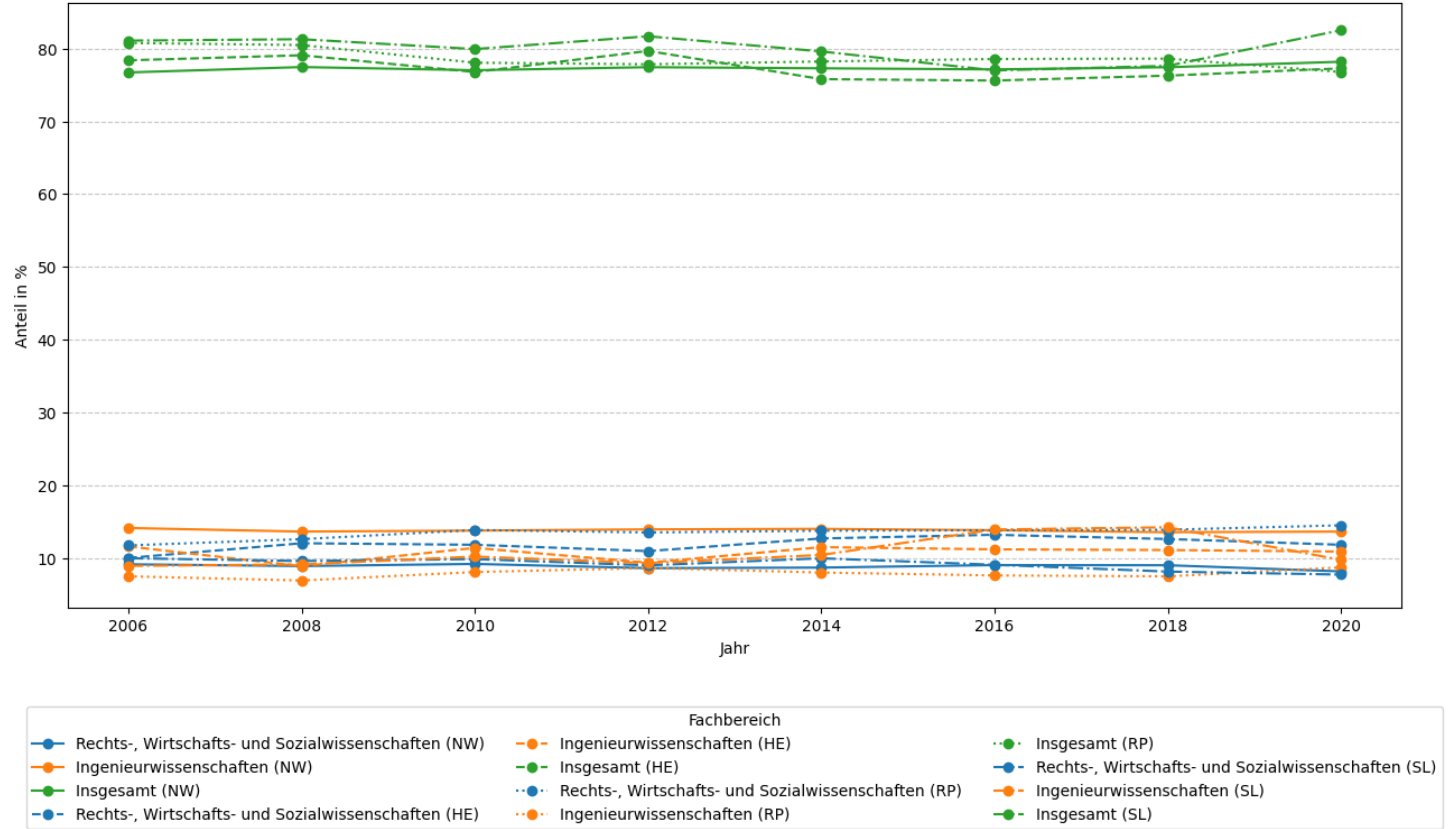
Anteile der Fachbereiche an den Gesamtausgaben im Osten (SN, TH, ST) von 2006-2020



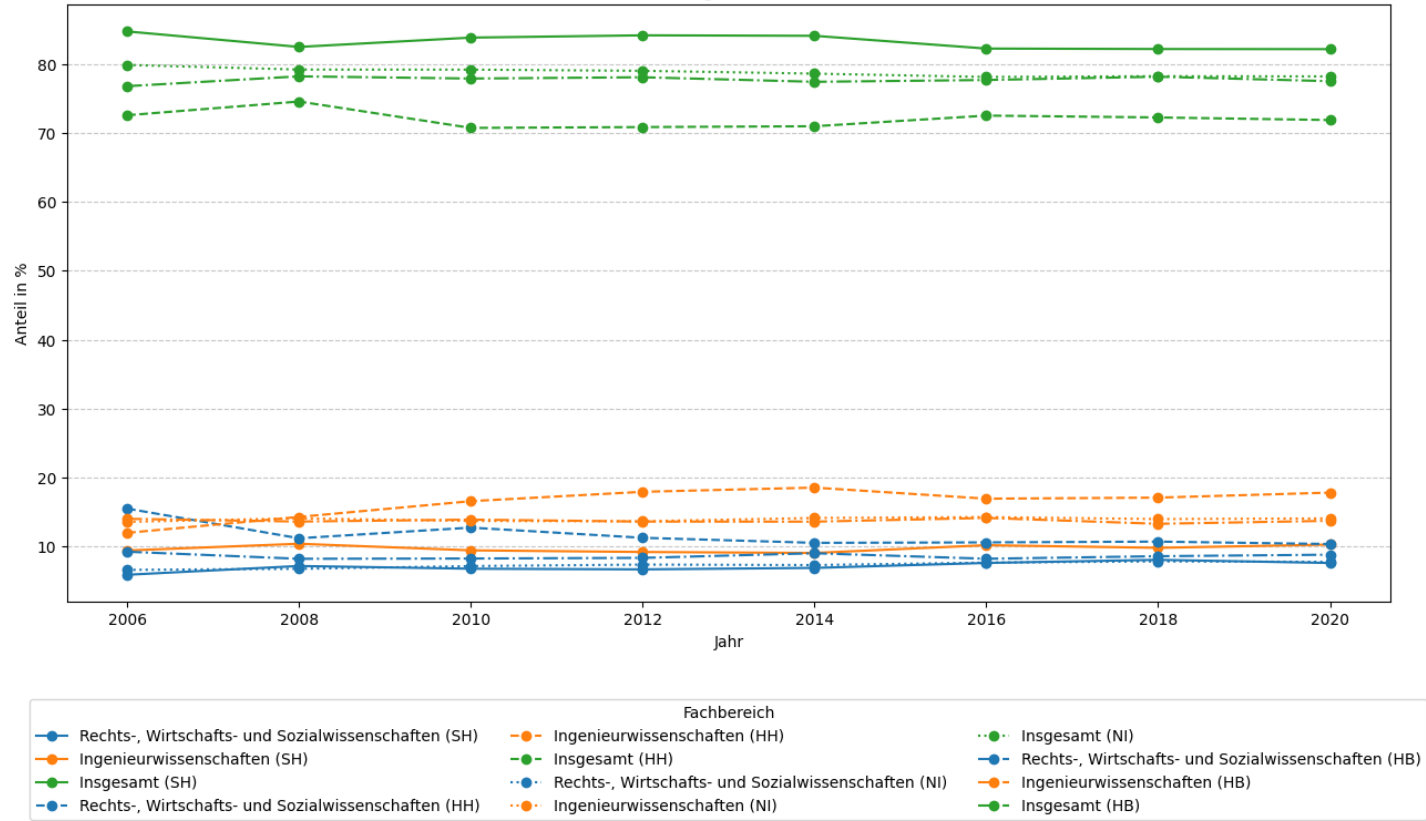
Anteile der Fachbereiche an den Gesamtausgaben im Osten (BB, MV, BE) von 2006-2020



Anteile der Fachbereiche an den Gesamtausgaben im Westen (NW, HE, RP, SL) von 2006-2020



Anteile der Fachbereiche an den Gesamtausgaben im Norden (SH, HH, NI, HB) von 2006-2020



Aufgabe 4: Zeitliche Entwicklung der Studierendenzahlen, der Personalzahlen und der Ausgaben



Wie haben sich die absoluten Zahlen der Studierenden und des Hochschulpersonals in den Fachbereichen Betriebswirtschaftslehre (BWL)/Wirtschaftswissenschaften und Elektrotechnik sowie die Gesamtausgaben der Hochschulen in Deutschland in den Jahren 1998 bis 2020 geändert?

Aufgabe 4: Code

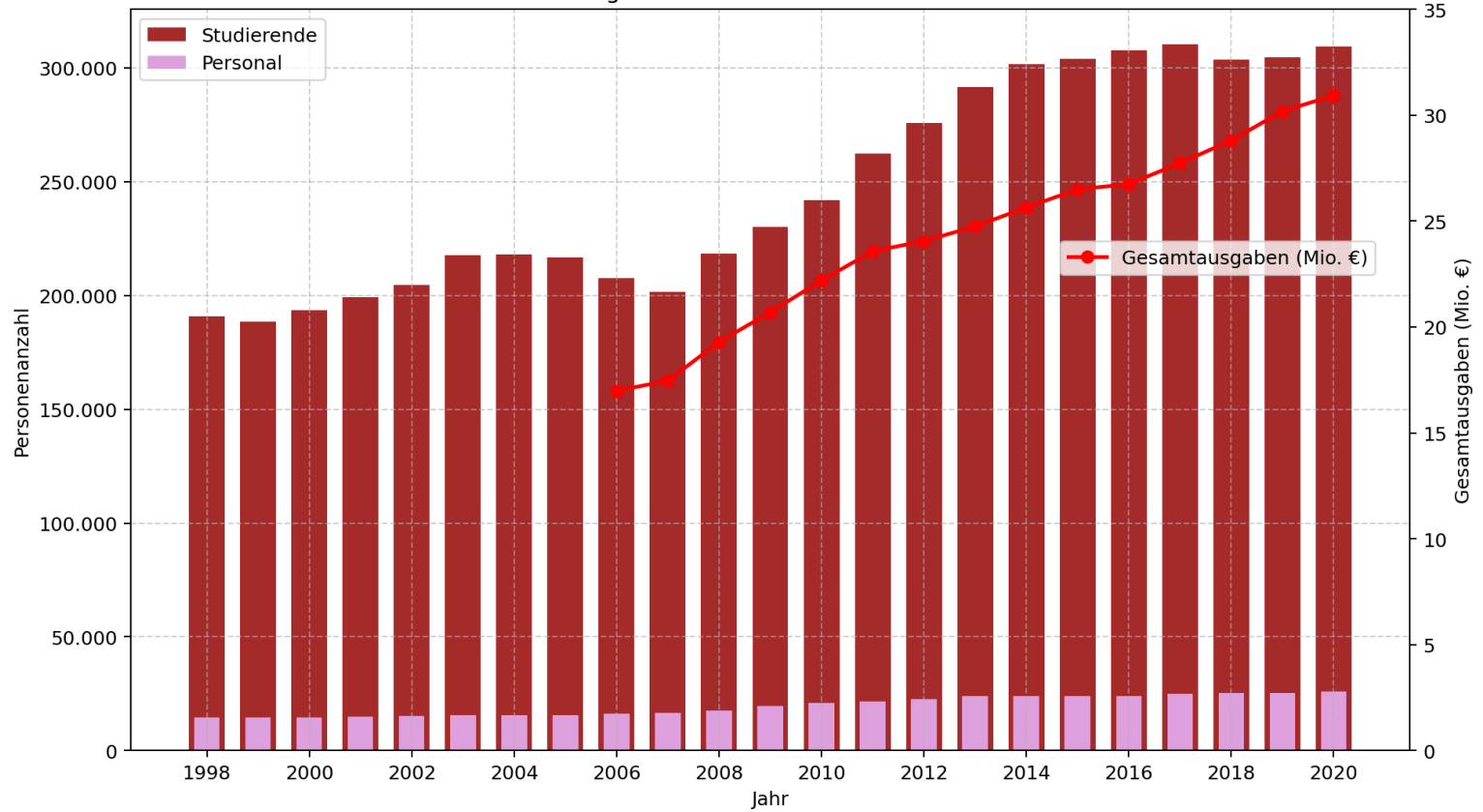
```
# Studierendenzahlen aggregieren (Summierung aller Studiengänge pro Jahr)
studierende_grouped = studierende.groupby('Jahr', as_index=False)['Insgesamt'].sum() # Gruppieren
studierende_grouped.rename(columns={'Insgesamt': 'Studierende'}, inplace=True) # Umbenennen der Spalte

# Personal-Daten aggregieren (nur Gesamtwerte für alle Fachbereiche summieren)
personal_filtered = personal[personal['Geschlecht'] == 'insgesamt'] # Filtern der Zeilen mit dem Wert 'insgesamt'
personal_melted = pd.melt(
    personal_filtered,
    id_vars=["Fachbereich", "Geschlecht"], # Beibehaltung dieser Spalten
    var_name="Jahr", # Umwandlung der Jahreswerte in eine eigene Spalte
    value_name="Personal" # Benennung der Werte in 'Personal'
)
personal_melted['Jahr'] = personal_melted['Jahr'].astype(str) # Sicherstellen, dass Jahr als String vorliegt
personal_grouped = personal_melted.groupby('Jahr', as_index=False)['Personal'].sum() # Summieren der Personal-Werte pro Jahr

# Gesamtausgaben berechnen (Summierung aller Bundesländer für jedes Jahr)
ausgaben_numeric = ausgaben.drop(columns=['Jahr', 'Fachbereich']) # Entfernen nicht-numerischer Spalten
ausgaben['Gesamtausgaben'] = ausgaben_numeric.sum(axis=1) # Berechnung der Gesamtausgaben durch Summieren der Spalten
ausgaben_grouped = ausgaben[['Jahr', 'Gesamtausgaben']].groupby('Jahr', as_index=False).sum()

# Zusammenführen der DataFrames nach Jahr
merged = pd.merge(studierende_grouped, personal_grouped, on='Jahr', how='outer')
merged = pd.merge(merged, ausgaben_grouped, on='Jahr', how='outer') # Hinzufügen der Gesamtausgaben
merged['Jahr'] = merged['Jahr'].astype(int) # Umwandlung der Jahr-Spalte in Integer
merged = merged.sort_values(by='Jahr') # Sortieren der Daten nach Jahr
```

Entwicklung von Studierendenzahlen und Personalzahlen (in den Studiengängen Betriebswirtschaftslehre und Elektrotechnik) sowie Ausgaben der Hochschulen in Deutschland



Quellen

Statistisches Bundesamt. (2024, 17. Januar). *Mehr als ein Drittel der Studienanfängerinnen und -anfänger im MINT-Bereich sind Frauen* [Pressemeldung]. https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/01/PD24_N003_213.html

Hochschulrektorenkonferenz. (2020, Oktober). *Statistische Daten zu Studienangeboten an Hochschulen in Deutschland Studiengänge, Studierende, Absolventinnen und Absolventen. Statistiken zur Hochschulpolitik 1/2020*. https://www.hrk.de/fileadmin/redaktion/hrk/02-Dokumente/02-03-Studium/02-03-01-Studium-Studienreform/HRK_Statistik_BA_MA_UEbrige_WiSe_2020_21_finale.pdf

Komm, mach MINT. *Was ist Komm, mach MINT?*. <https://www.komm-mach-mint.de/komm-mach-mint>.

Erhardt-Maciejewski, Christian (2021, 24. November): *Städte-Ranking: Deutschlands erfolgreichste Industriestädte*. <https://kommunal.de/ranking-industriestaedte-2021>. (16.02.2025)

Quellen

Statistisches Bundesamt (2025, 6. Februar): *Ausgaben der Hochschulen: Bundesländer, Jahre, Hochschulart, Fächergruppen*. <https://www-genesis.destatis.de/datenbank/online/statistic/21371/table/21371-0002/search/s/MjEzNzE=> (16.02.2025).