

In [369...

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

from nutil.plot import paperStyle # (pip install
git+https://github.com/anki-xyz/nutil)
```

Further information about the DataFrame used today:

<https://github.com/fivethirtyeight/data/tree/master/college-majors>

In [3]:

```
download_url =
("https://raw.githubusercontent.com/fivethirtyeight/data/master/college-
majors/recent-grads.csv")
```

In [370...

```
a = pd.read_csv(download_url)
```

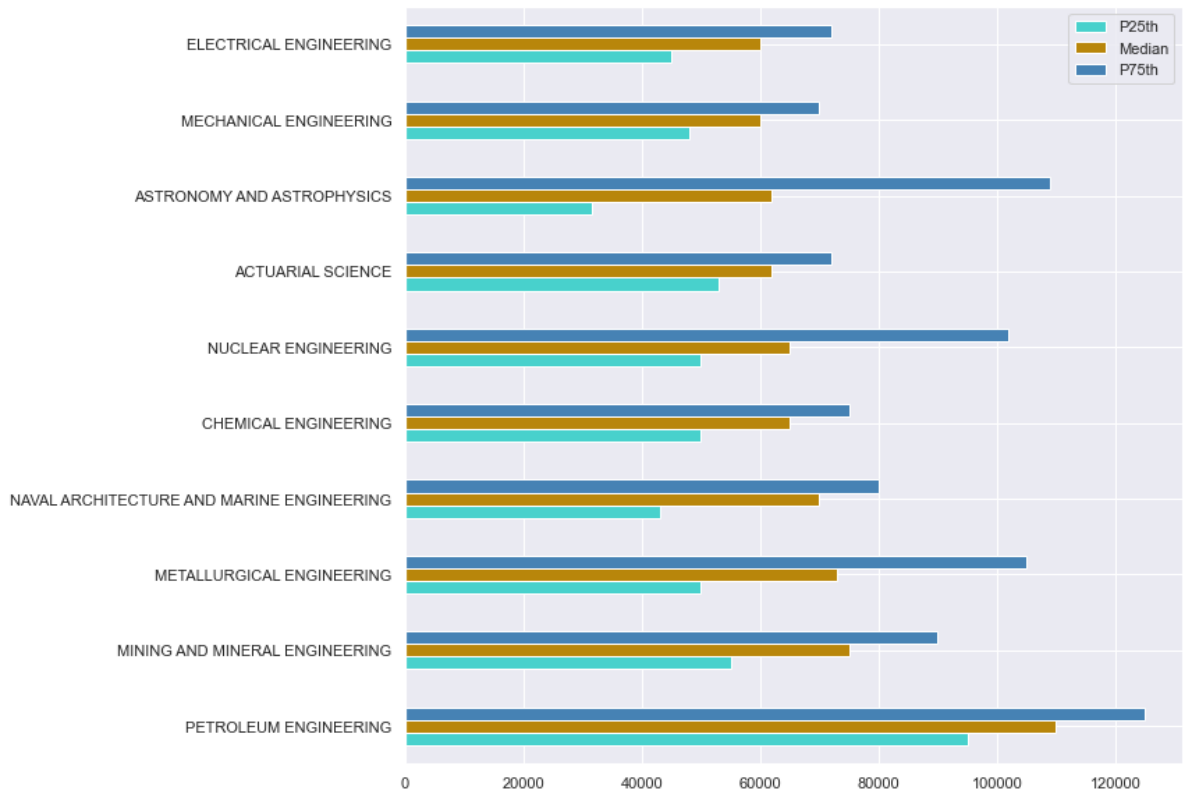
Task a)

Plot the top 10 majors depending on the median earnings of full-time, year-round workers. Make it nice looking and prepare the plot for publication. Then save it in a suitable file format to load it into Inkscape.

Median	Median earnings of full-time, year-round workers
P25th	25th percentile of earnings
P75th	75th percentile of earnings

In [388...

```
a.sort_values(by = ['Median'], ascending=False)
a.iloc[:10].plot(x="Major", y=["P25th", "Median", "P75th"],
kind="barh",figsize=(10,10),color=
('mediumturquoise','darkgoldenrod','steelblue'))
plt.ylabel('')
plt.show()
```



Task b)

Plot the boxplots of median earnings for the top 5 categories (as measured by median earnings). Then add the raw sample points to the plot using a suitable plot type. Make it nice looking and prepare the plot for publication. Then save it in a suitable file format to load it into Inkscape.

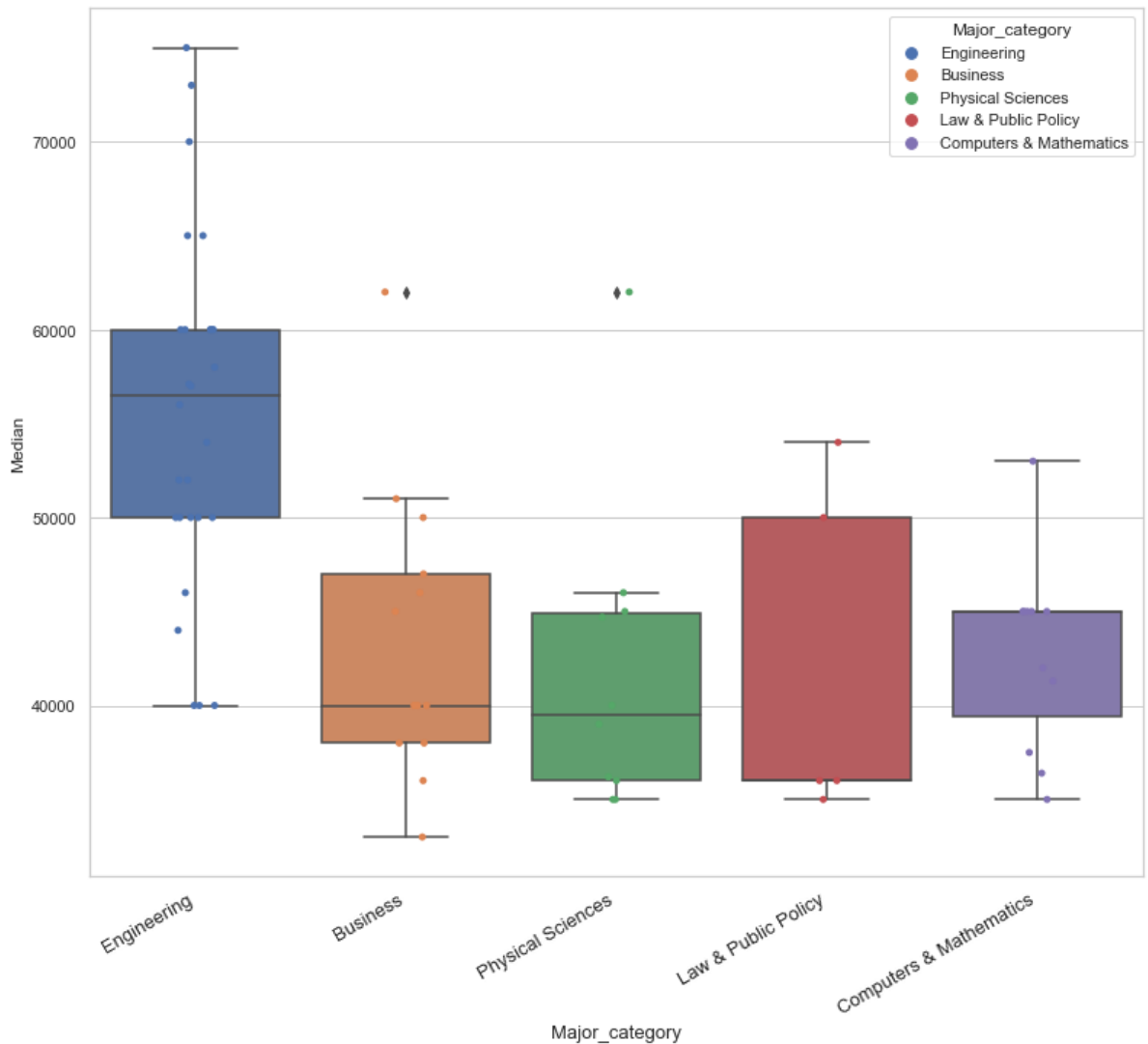
In [284...

```
b = pd.read_csv("new.csv")
```

```
# Q3 + whis*(Q3 - Q1) u = 56250 + 1.5*(56250 - 40000) # Q1 - whis*(Q3 - Q1) l = 40000 - 1.5*(56250 - 40000)
# Apply limits to DataFrame b = b[(b['Median'] <= u) & (b['Median'] >= l)]
```

In [385...

```
sns.set_style('whitegrid')
c = sns.boxplot(data=b, x='Major_category', y='Median')
c = sns.stripplot(x="Major_category",
y="Median",data=b,hue="Major_category")
c.set_xticklabels(
    c.get_xticklabels(),
    rotation=30,
    horizontalalignment='right',
    fontsize='13',
)
c.set_xlabel('Major_category',fontsize='13')
sns.set(rc={'figure.figsize':(13,11)})
plt.show()
```



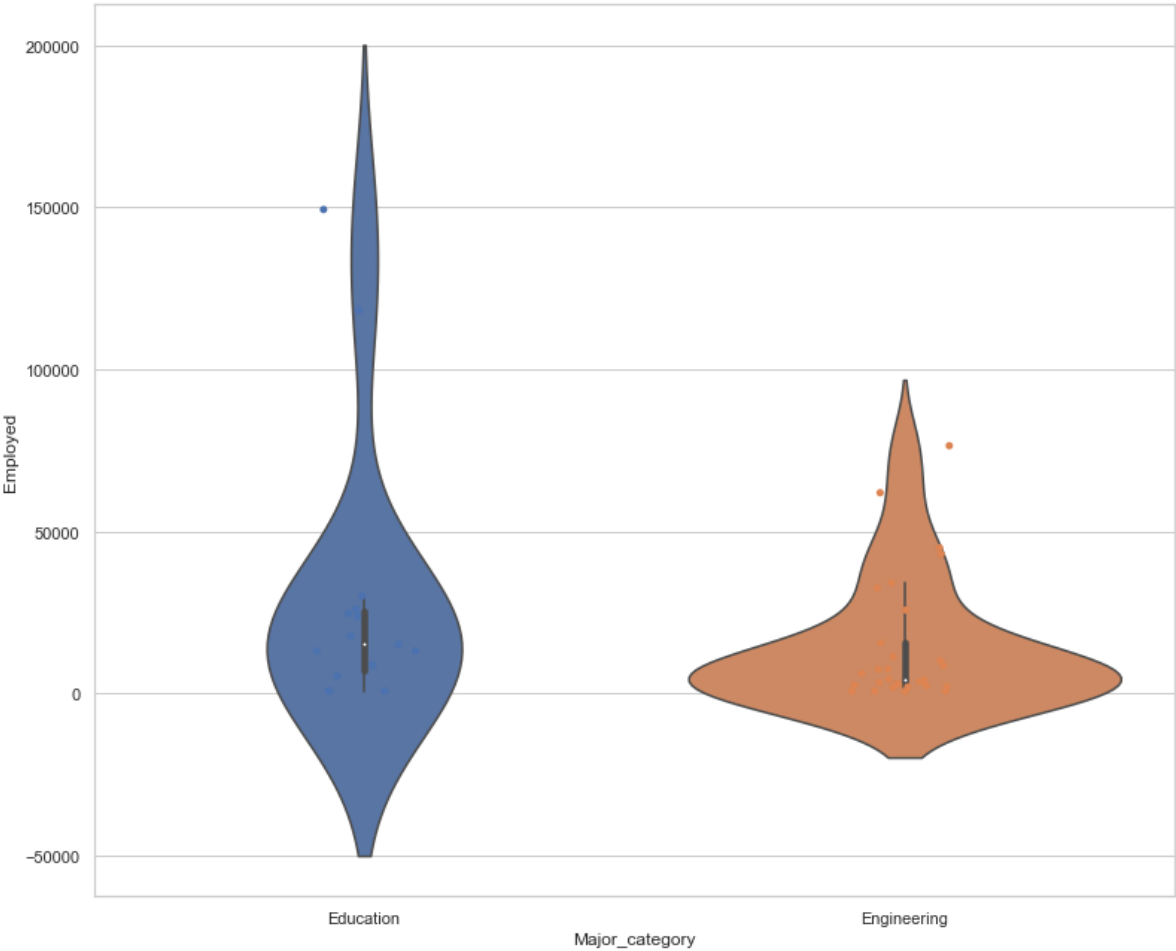
Task c)

Now it's your turn to create a plot yourself. Choose one of the following plots along with the corresponding data (from the DataFrame above) and make it look nice: Lineplot, Violinplot or Scatterplot. Make it nice looking and prepare the plot for publication. Then save it in a suitable file format to load it into Inkscape.

In [8]: `# todo: your solution here`

In [381...]: `a1 = a.sort_values(by = ['Major_category'], ascending=True)
a1 = a1.iloc[61:105, :]`

In [382...]: `sns.set_style('whitegrid')
ax = sns.violinplot(x='Major_category', y='Employed', data=a1)
ax = sns.stripplot(x="Major_category", y="Employed", data=a1)`



In []: