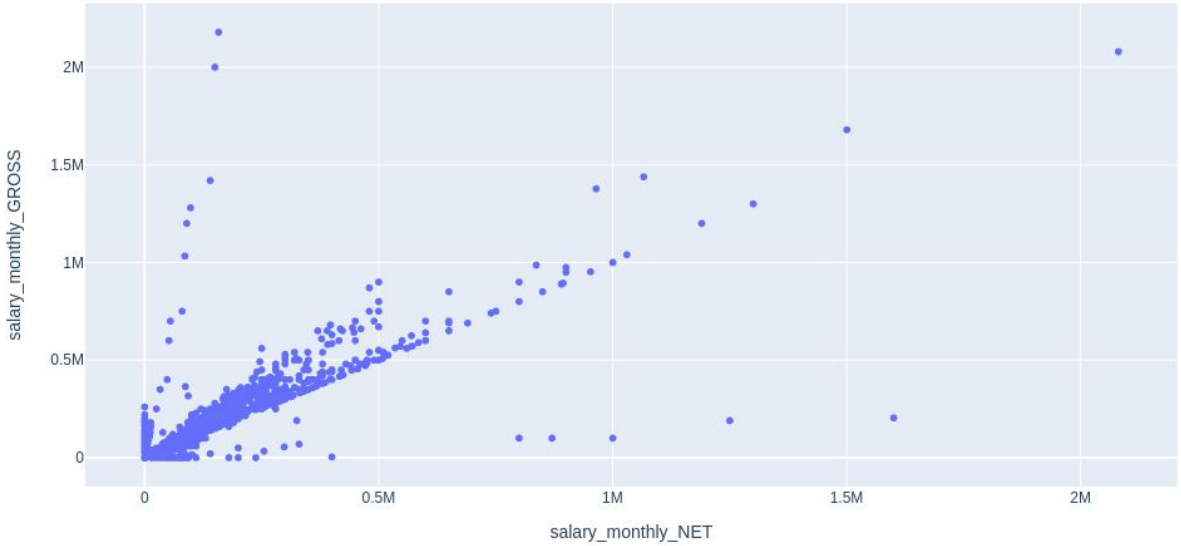
Data Visualization

Revision Class 02 - Two random variables

DigitalLab@LaPlataforme_

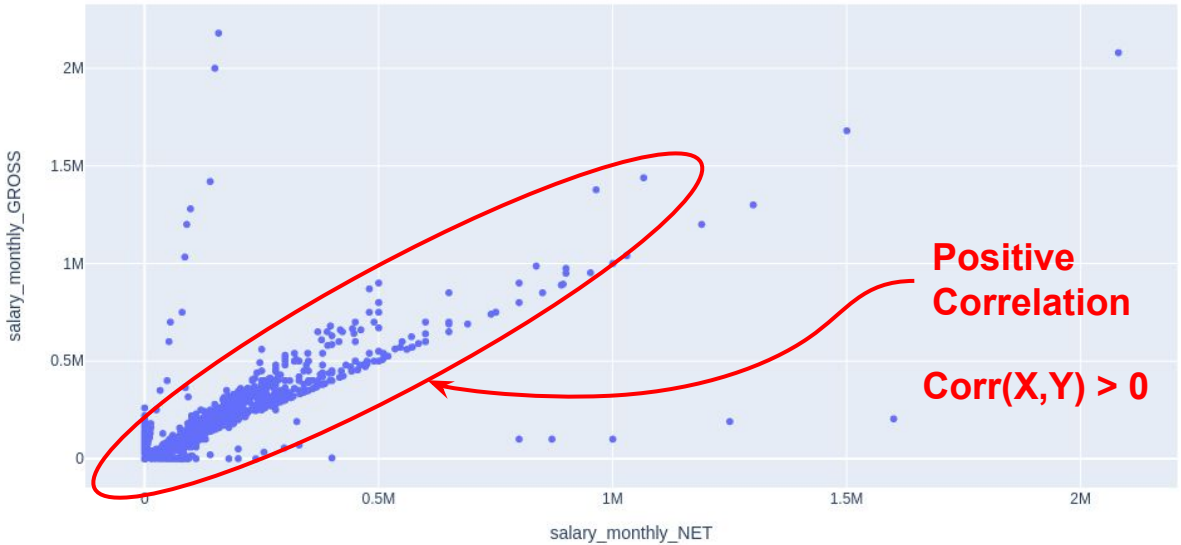
Two variables: Scatterplots

The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.



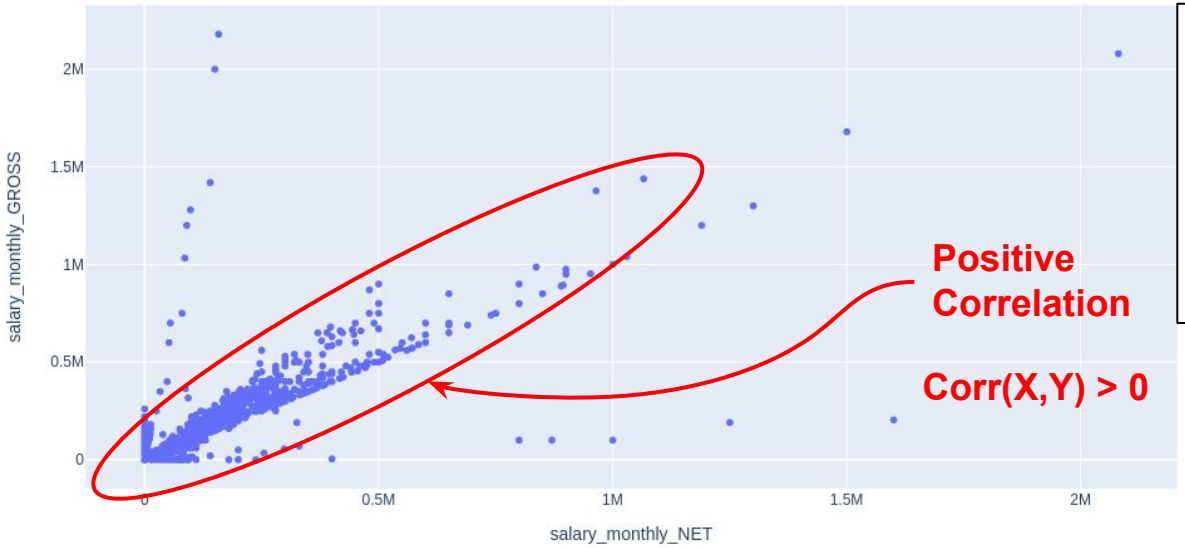
Two variables: Scatterplots

The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.



Two variables: Scatterplots

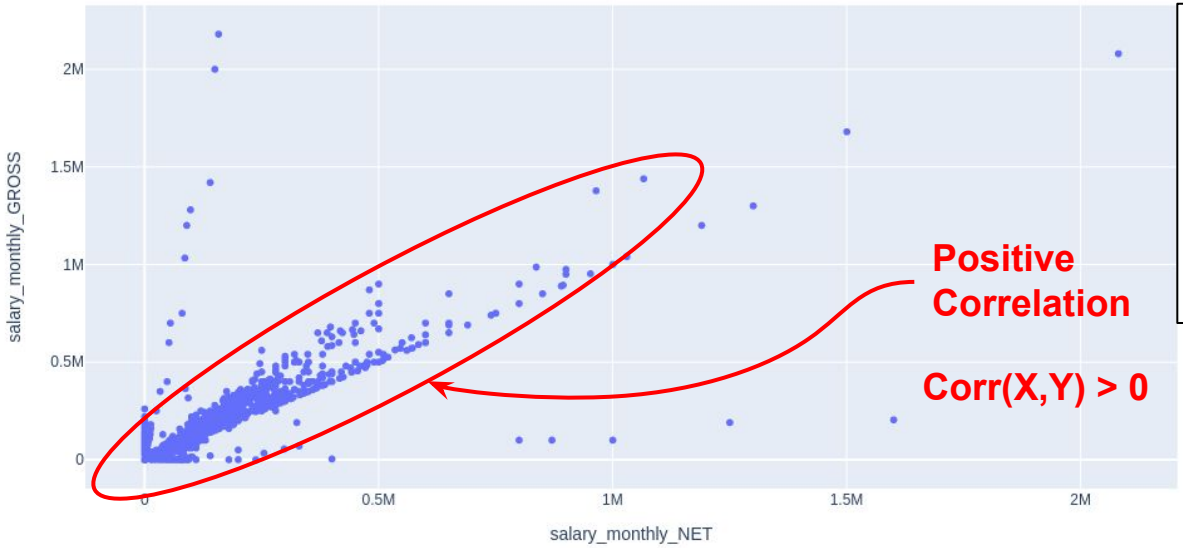
The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.




```
fig = px.scatter(
     df,
      y="salary_monthly_GROSS",
      x="salary_monthly_NET")
fig.show()
```

Two variables: Scatterplots

The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.

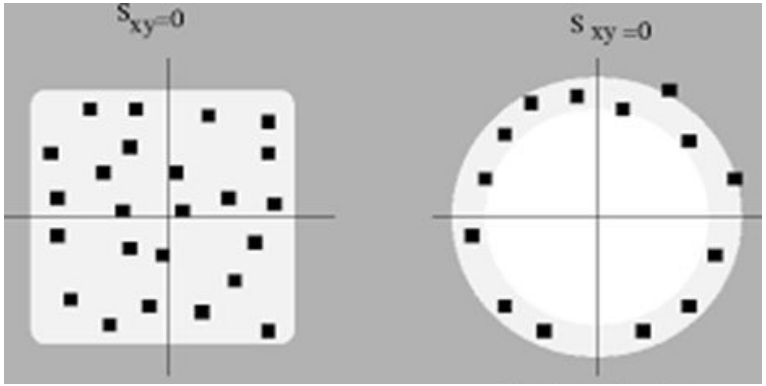



```
fig = px.scatter(
     df,
      y="salary_monthly_GROSS",
      x="salary_monthly_NET")
fig.show()
```

IMPORTANT! We can only use this plot with numerical two continuous r.v.

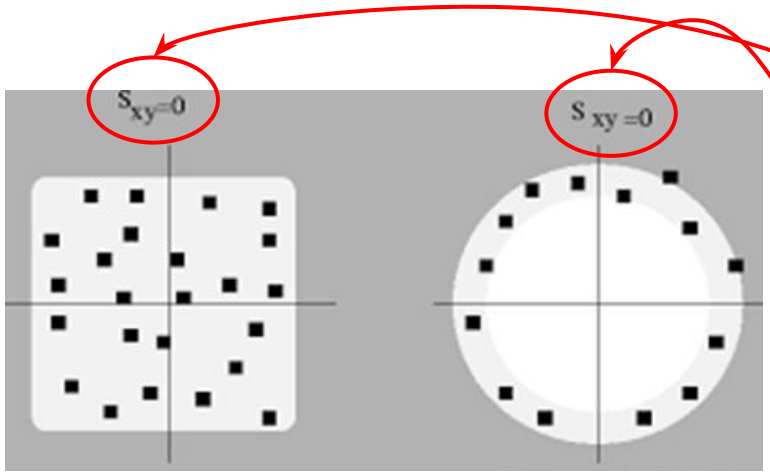
Two variables: Scatterplots

The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.



Two variables: Scatterplots

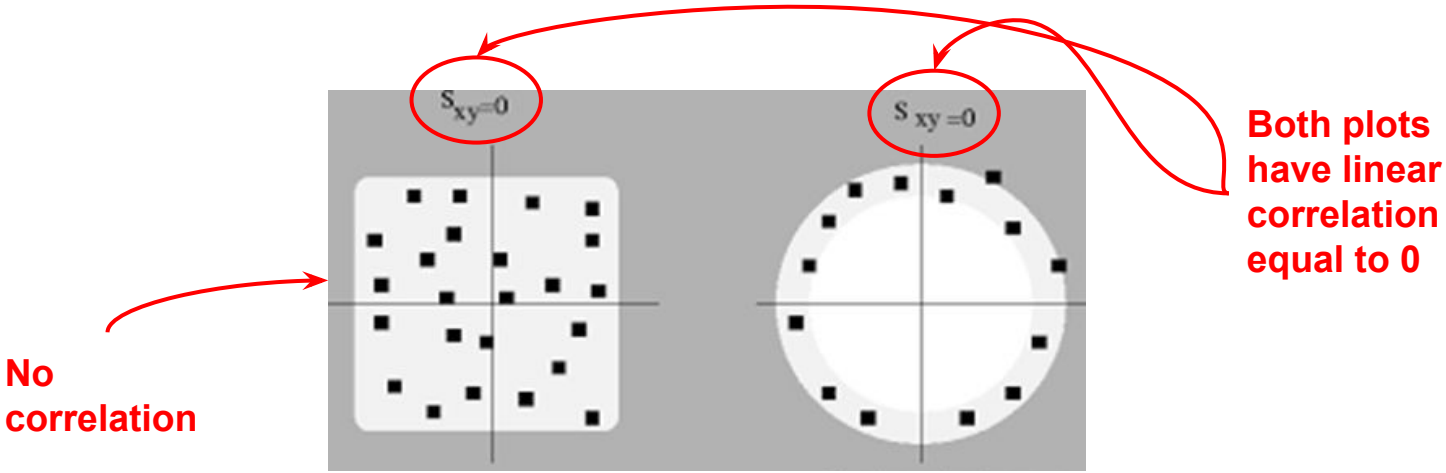
The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.



Both plots have linear correlation equal to 0

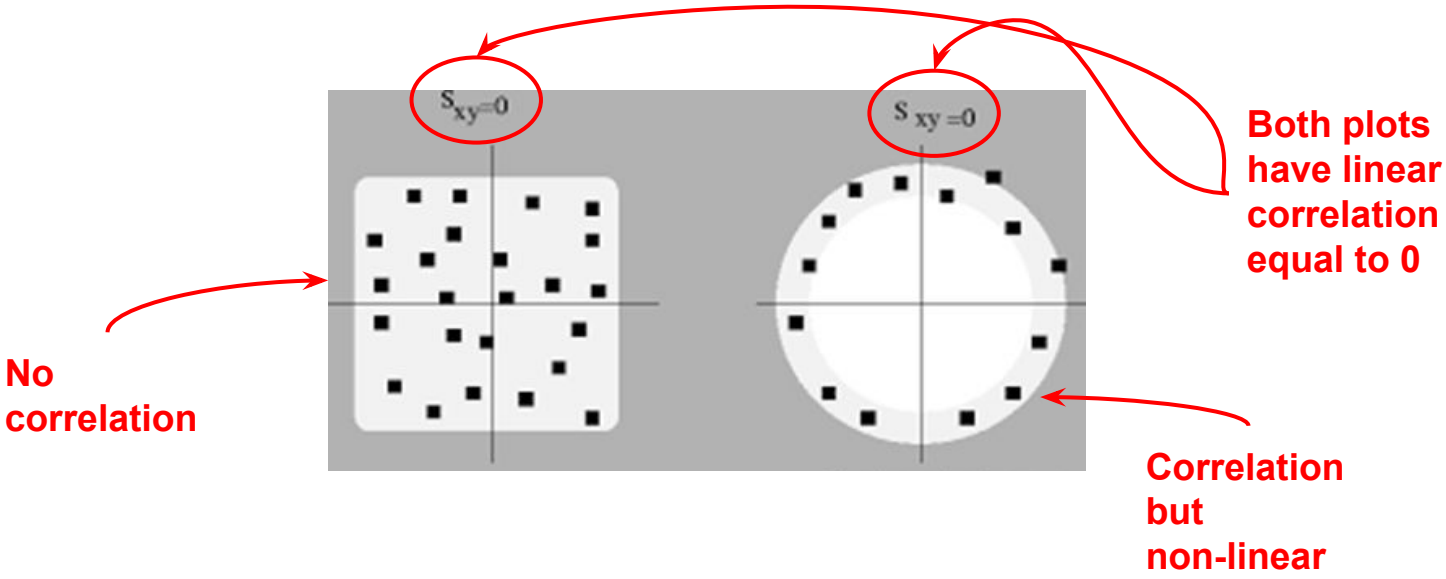
Two variables: Scatterplots

The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.



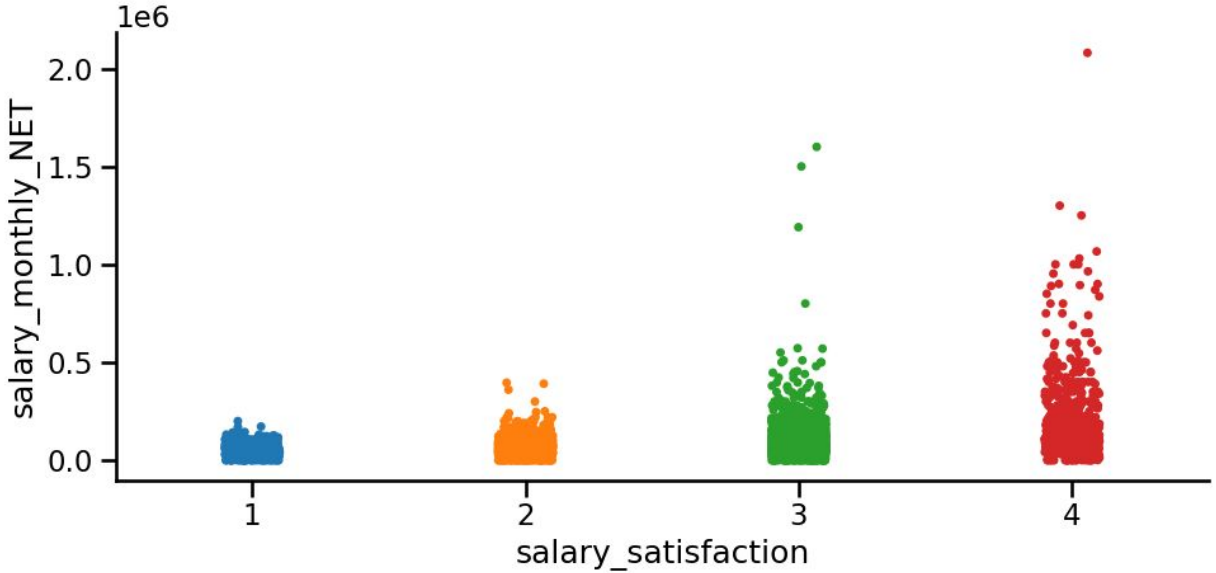
Two variables: Scatterplots

The scatterplot shows the **relationship between two numerical r.v.** X and Y by mapping realizations of both variables into an 2D plot.



Two variables: Catplots

The catplots shows the **relationship between one categorical r.v.** X and **one numerical** Y.



Two variables: Catplots

The catplots shows the **relationship between one categorical r.v.** X and **one numerical** Y.

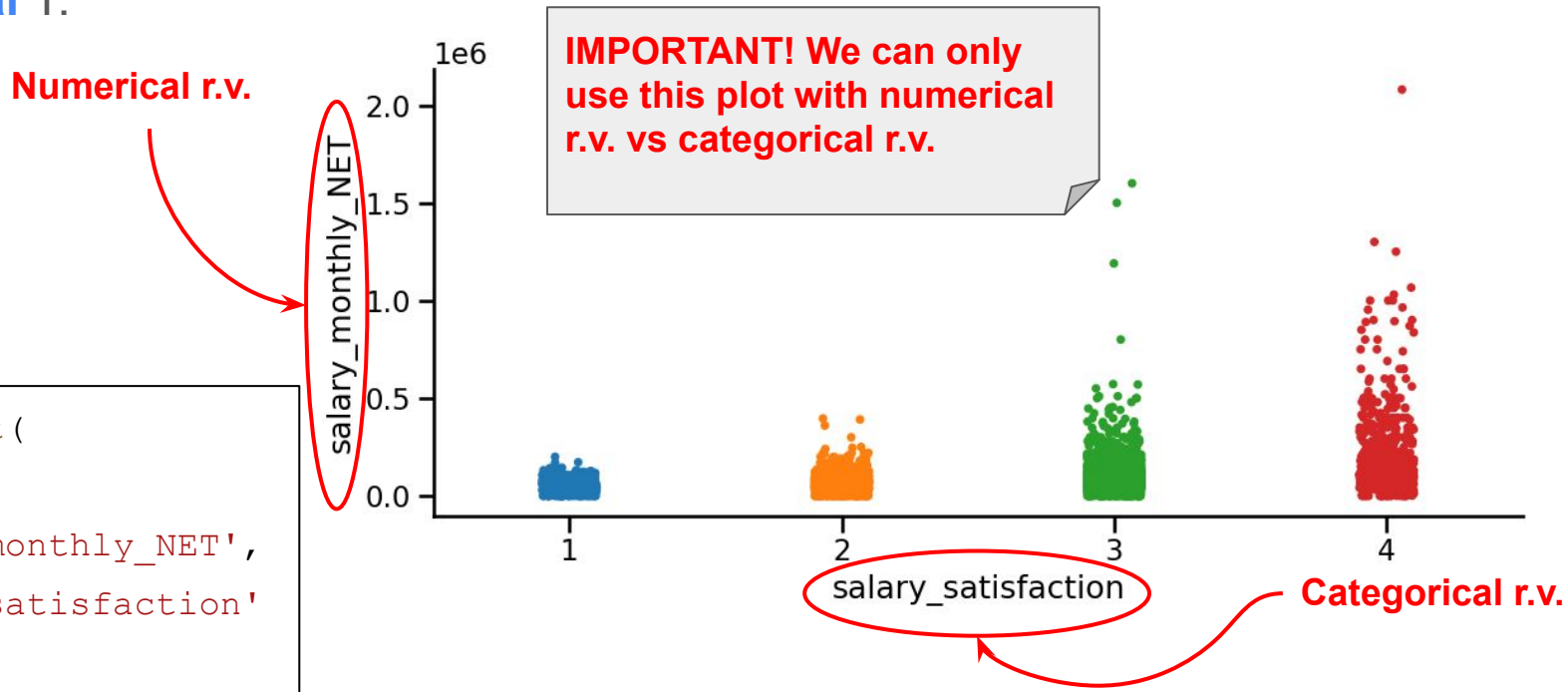
Two variables: Catplots

The catplots shows the **relationship between one categorical r.v.** X and **one numerical** Y.


```
sns.catplot(
data=df,
y='salary_monthly_NET',
x='salary_satisfaction'
)
```

Two variables: Catplots

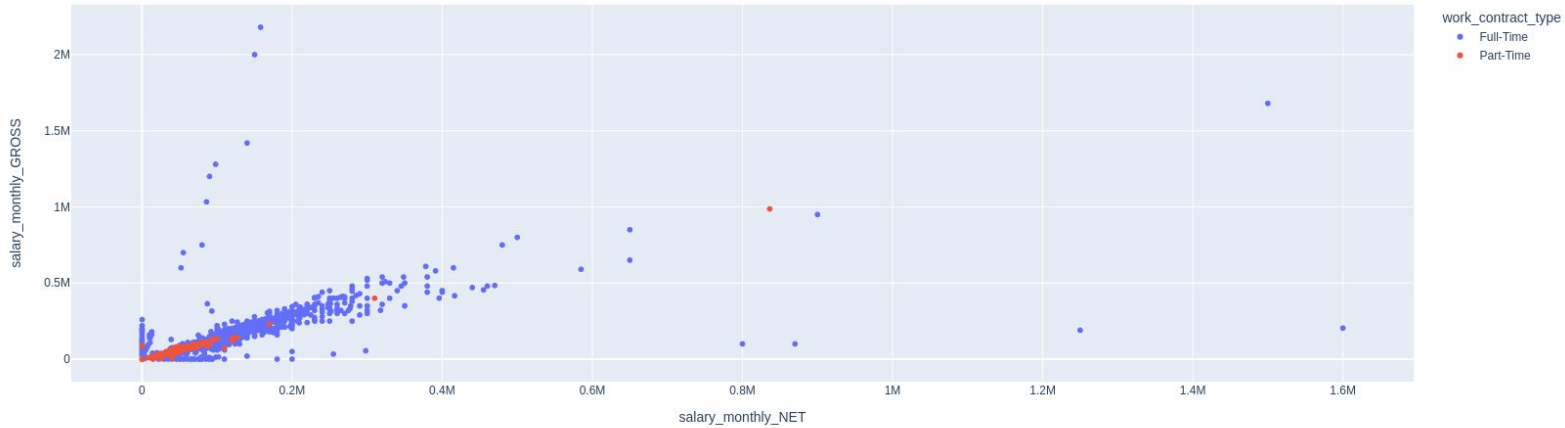
The catplots shows the **relationship between one categorical r.v.** X and **one numerical** Y.




```
sns.catplot(
data=df,
y='salary_monthly_NET',
x='salary_satisfaction'
)
```

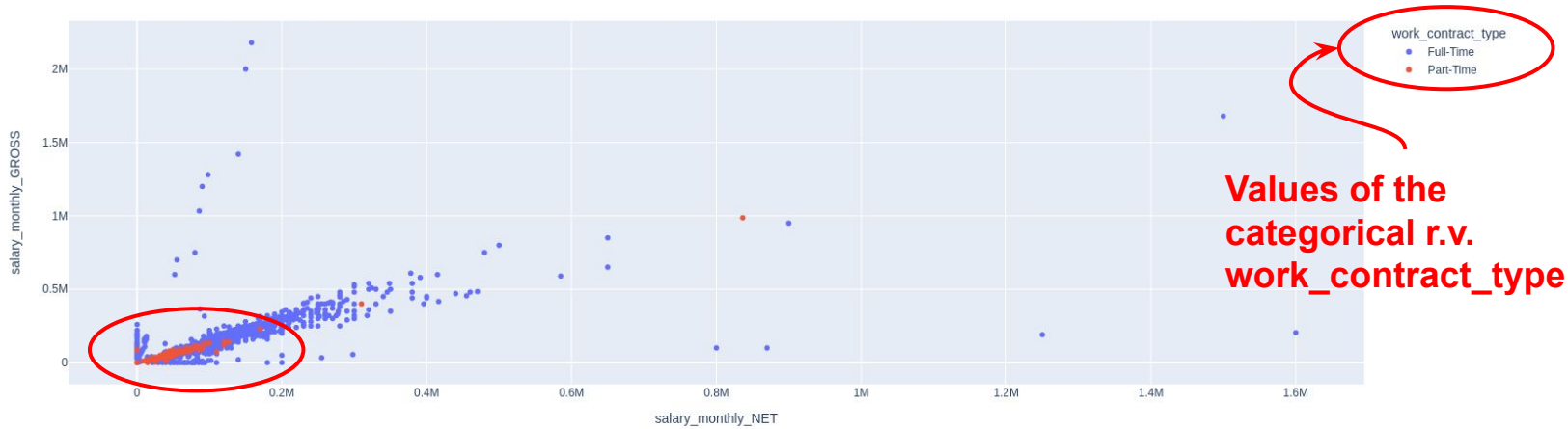
The power of hue

It controls when you want to **visualize different subsets of the data**. It can be used with **only categorical variables** and it is preferable where the **number of categories is small**.



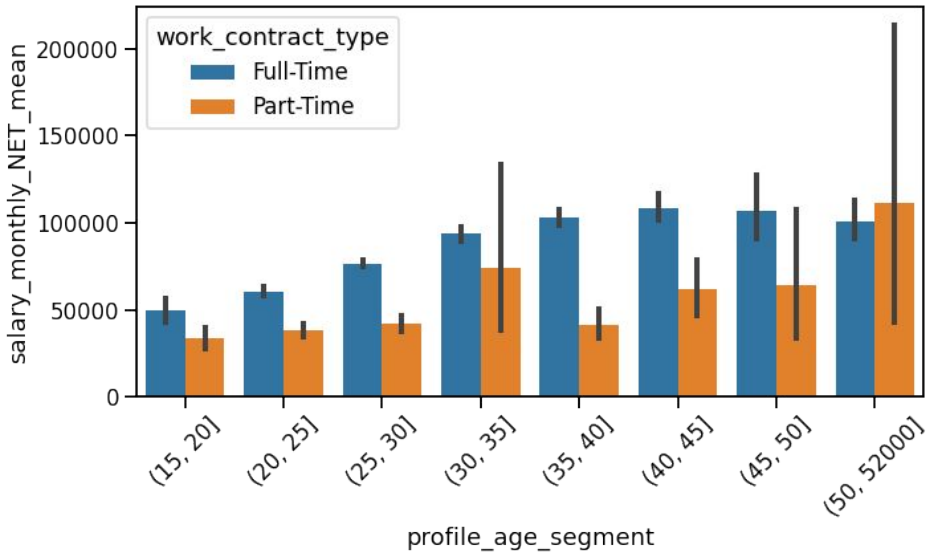
The power of hue

It controls when you want to **visualize different subsets of the data**. It can be used with **only categorical variables** and it is preferable where the **number of categories is small**.



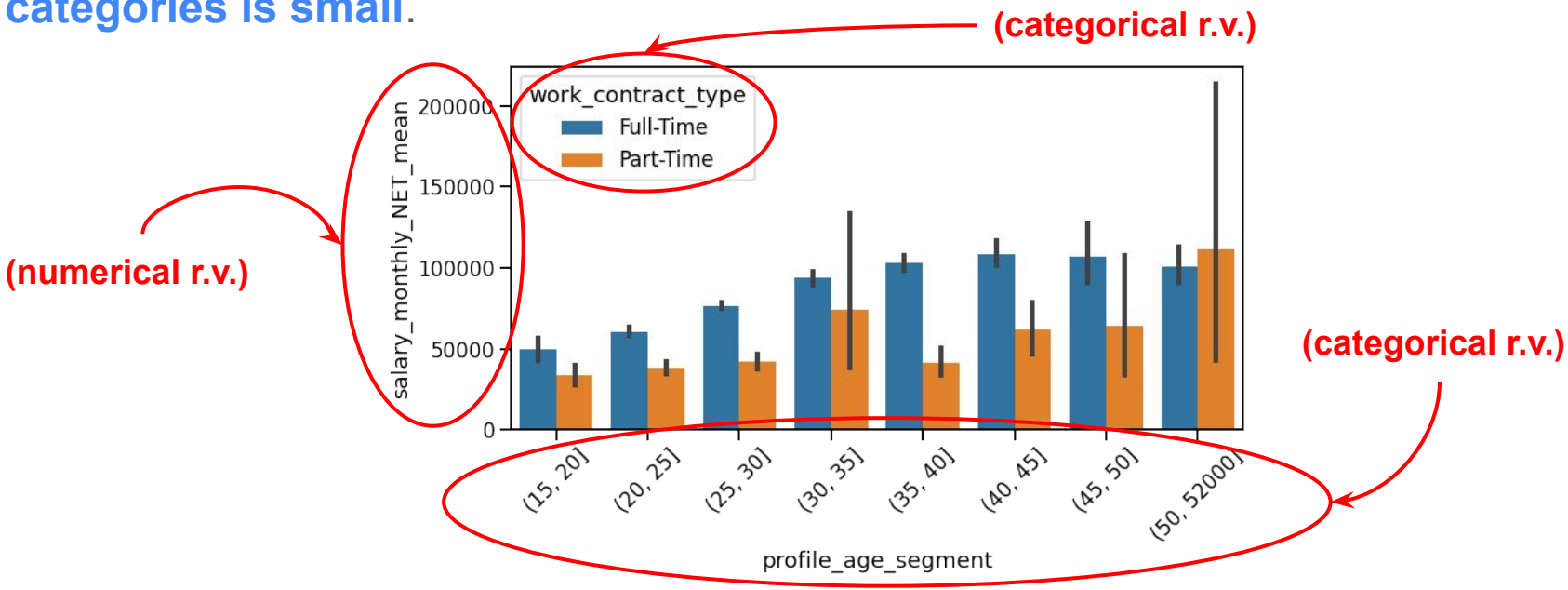
The power of hue

It controls when you want to **visualize different subsets of the data**. It can be used with **only categorical variables** and it is preferable where the **number of categories is small**.



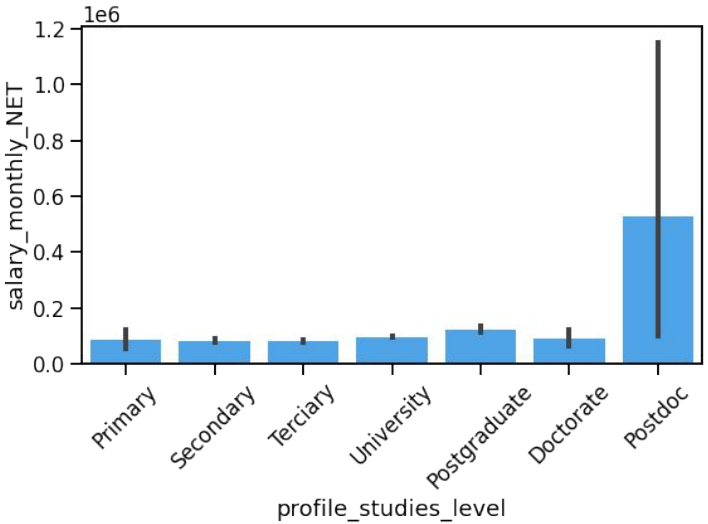
The power of hue

It controls when you want to **visualize different subsets of the data**. It can be used with **only categorical variables** and it is preferable where the **number of categories is small**.



From Seaborn to Plotly

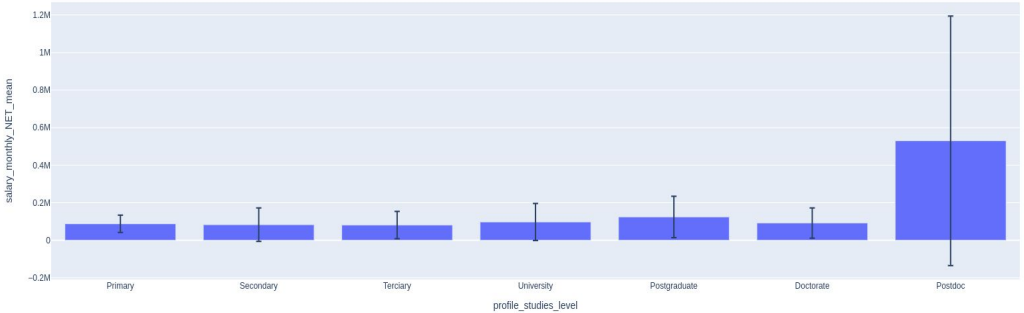
The plotly library is an interactive open-source plotting library that supports the creation of **more personalized plots than seaborn** but at the same time with a harder learning curve.




```
seaborn.barplot(
data=df,
y="salary_monthly_NET",
x='profile_studies_level',
estimator=numpy.mean,
ci=95)
```

From Seaborn to Plotly

The plotly library is an interactive open-source plotting library that supports the creation of **more personalized plots than seaborn** but at the same time with a harder learning curve.

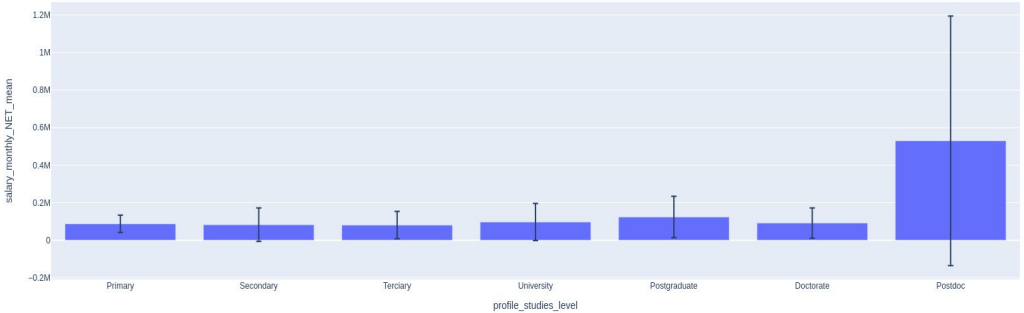



```
fig = px.bar(
df_studies_level_mean,
x='profile_studies_level',
y='salary_monthly_NET_mean',
error_y="salary_monthly_NET_std")
fig.show()
```

From Seaborn to Plotly

The plotly library is an interactive open-source plotting library that supports the creation of **more personalized plots than seaborn** but at the same time with a harder learning curve.

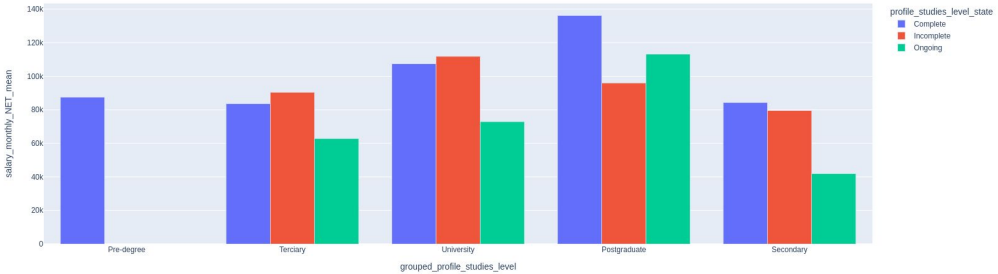
IMPORTANT! Sometimes we need to calculate the aggregation




```
fig = px.bar(
df_studies_level_mean,
x='profile_studies_level',
y='salary_monthly_NET_mean',
error_y="salary_monthly_NET_std")
fig.show()
```

From Seaborn to Plotly

The plotly library is an interactive open-source plotting library that supports the creation of **more personalized plots than seaborn** but at the same time with a harder learning curve.

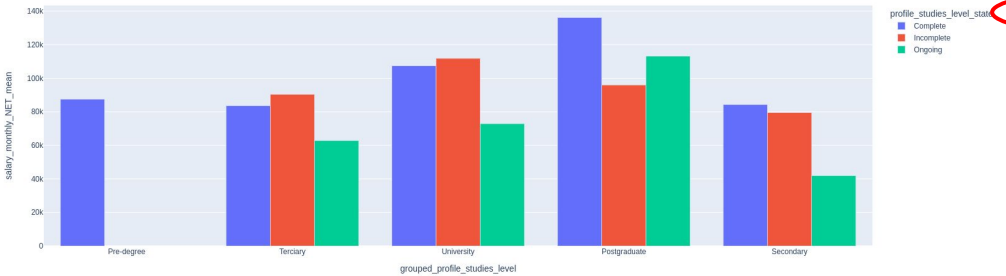



```
fig = px.bar(
df_grouped_studies_level_mean,
x='profile_studies_level',
y='salary_monthly_NET_mean',
color='profile_studies_level_state',
barmode='group')
fig.show()
```

From Seaborn to Plotly

The plotly library is an interactive open-source plotting library that supports the creation of **more personalized plots than seaborn** but at the same time with a harder learning curve.

Dataframe with the studies level, level state, and salary mean




```
fig = px.bar(
df_grouped_studies_level_mean,
x='profile_studies_level',
y='salary_monthly_NET_mean',
color='profile_studies_level_state',
barmode='group')
fig.show()
```