Introduction to categorical plots using Seaborn

WORKING WITH CATEGORICAL DATA IN PYTHON

Kasey Jones
Research Data Scientist

datacamp

Our third dataset

- Name: Las Vegas TripAdvisor Reviews - `reviews`
- Rows: 504
- Columns: 20

Las Vegas reviews

```
reviews.info()
```

```
RangeIndex: 504 entries, 0 to 503
Data columns (total 20 columns):
 #   Column              Non-Null Count  Dtype
     ------              --------------  -----
 0   User country        504 non-null    object
...
 6   Traveler type       504 non-null    object
 7   Pool                504 non-null    object
 8   Gym                 504 non-null    object
 9   Tennis court        504 non-null    object
...
dtypes: int64(7), object(13)
memory usage: 78.9+ KB
```

[1] https://www.kaggle.com/crawford/las-vegas-tripadvisor-reviews

Seaborn

- Introduction to Data Visualization with Seaborn
- Intermediate Data Visualization with Seaborn

Categorical plots:

```
import seaborn as sns
import matploblib.pyplot as plt

sns.catplot(...)
```

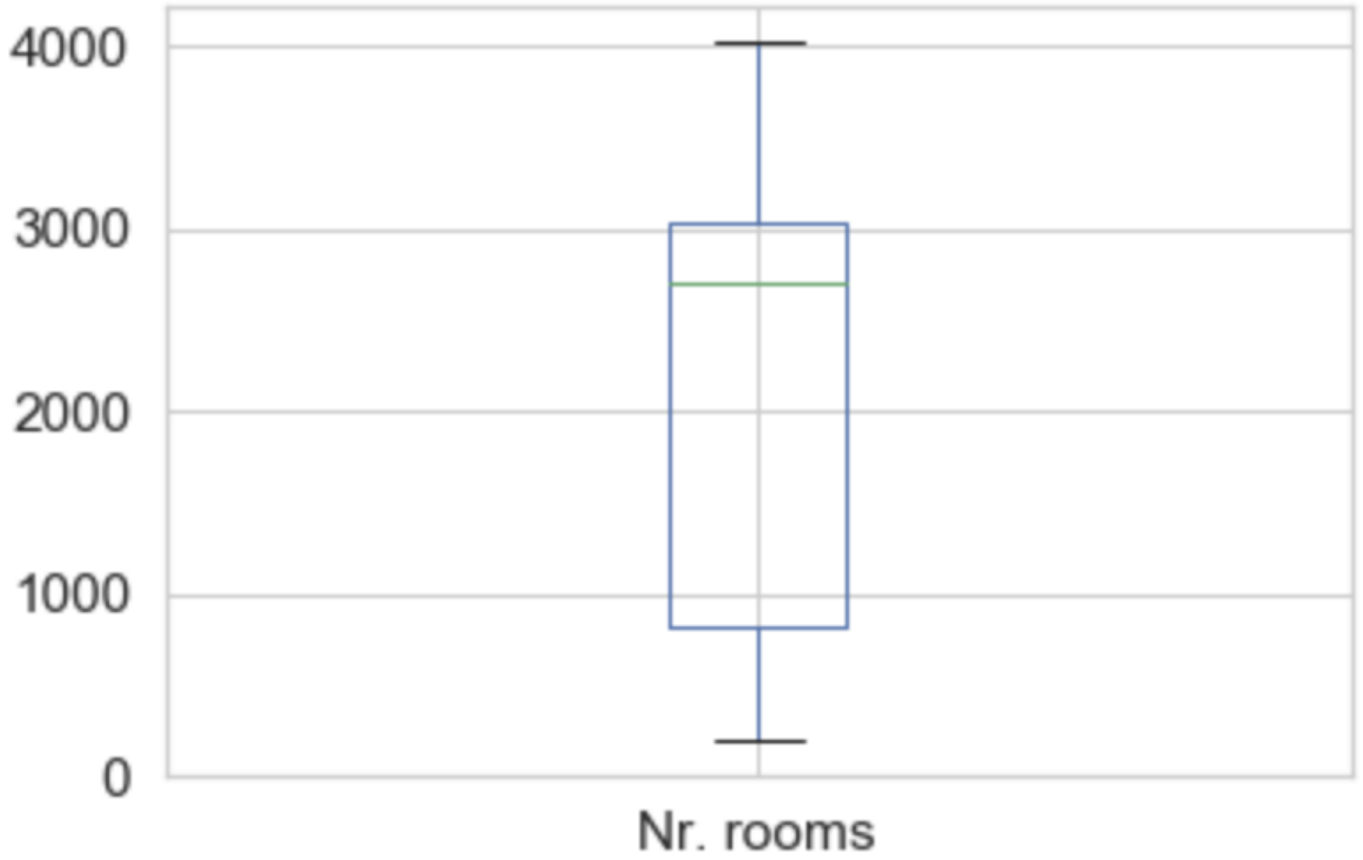
```
plt.show()
```

The catplot function

Parameters:

- `x` : name of variable in `data`
- `y` : name of variable in `data`
- `data` : a DataFrame
- `kind` : type of plot to create - one of: `"strip"` , `"swarm"` , `"box"` , `"violin"` , `"boxen"` , `"point"` , `"bar"` , or `"count"`

Box plot



Box plot wiki page

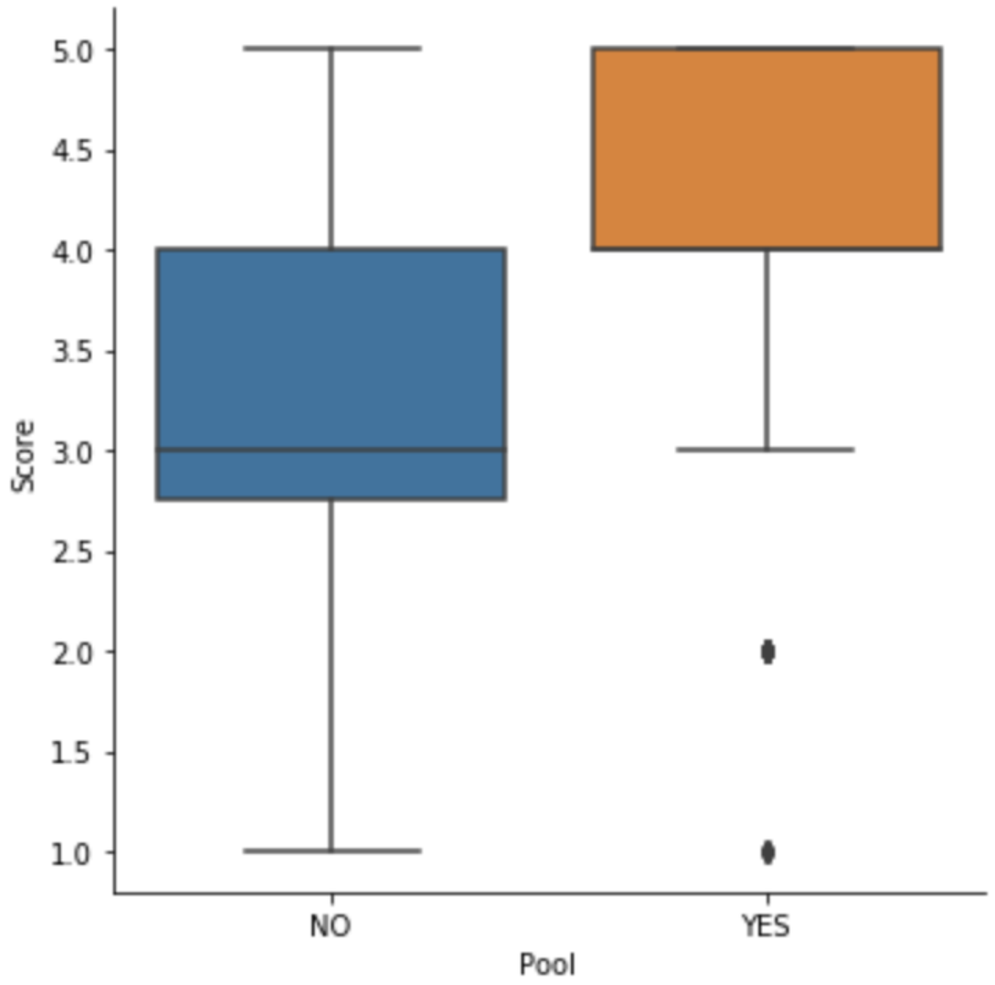
Review score

```
reviews["Score"].value_counts()
```

```
5    227
4    164
3     72
2     30
1     11
```

Box plot example

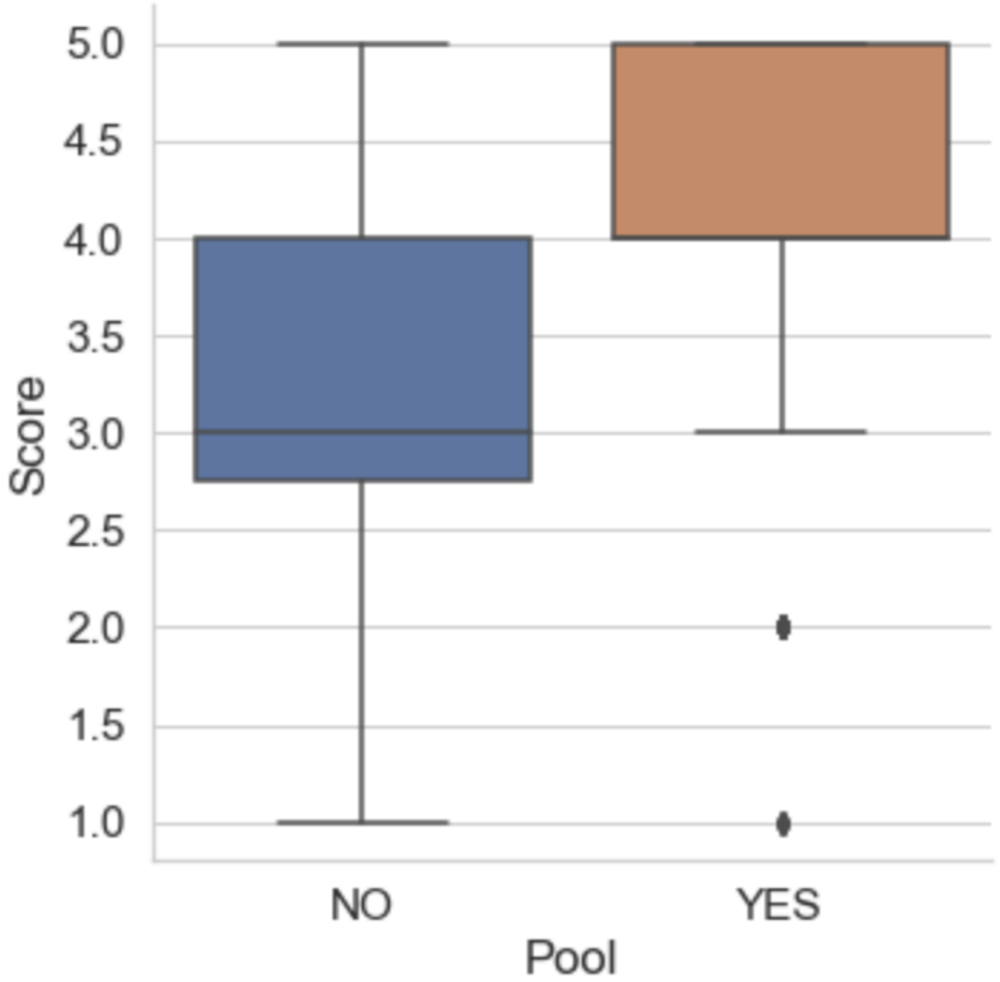
```
sns.catplot(
  x="Pool",
  y="Score",
  data=reviews,
  kind="box"
)
plt.show()
```



Two quick options

```
# Setting font size and plot background
sns.set(font_scale=1.4)
sns.set_style("whitegrid")
```

```
sns.catplot(
  x="Pool",
  y="Score",
  data=reviews,
  kind="box"
)
plt.show()
```



Boxplot practice

WORKING WITH CATEGORICAL DATA IN PYTHON

datacamp

Seaborn bar plots

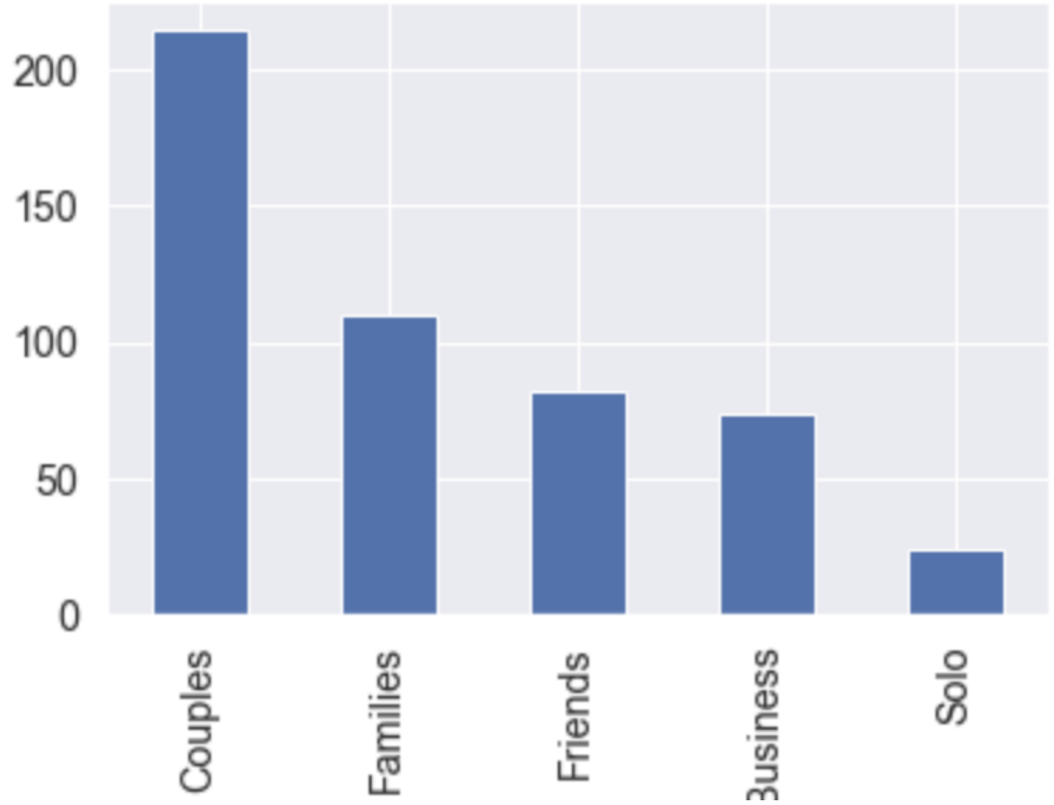
WORKING WITH CATEGORICAL DATA IN PYTHON

Kasey Jones
Research Data Scientist

datacamp

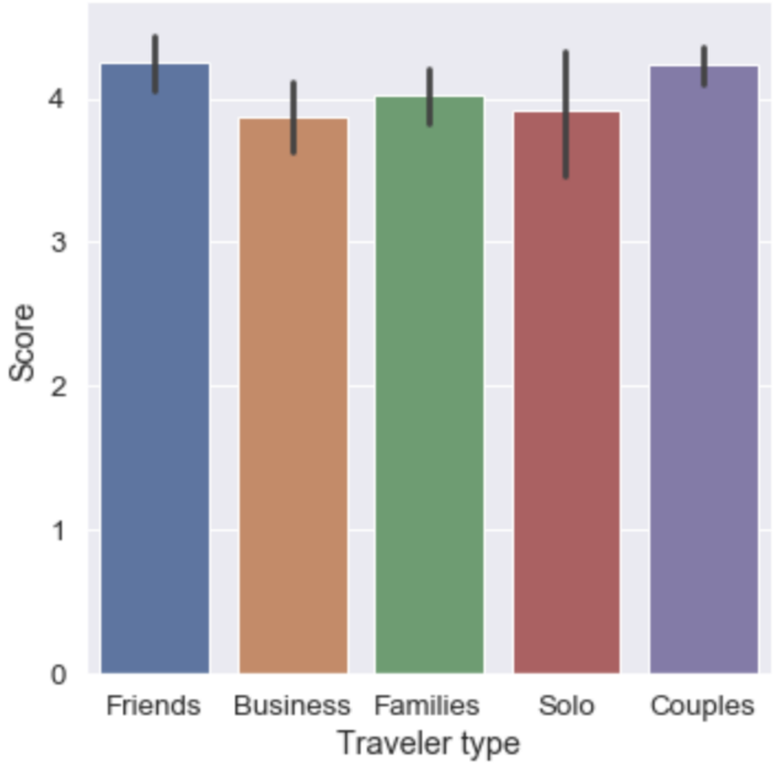
Traditional bar chart

```
# Code provided for clarity
reviews["Traveler type"].value_counts().plot.bar()
```



The syntax

```
sns.set(font_scale=1.3)
sns.set_style("darkgrid")
sns.catplot(x="Traveler type", y="Score", data=reviews, kind="bar")
```



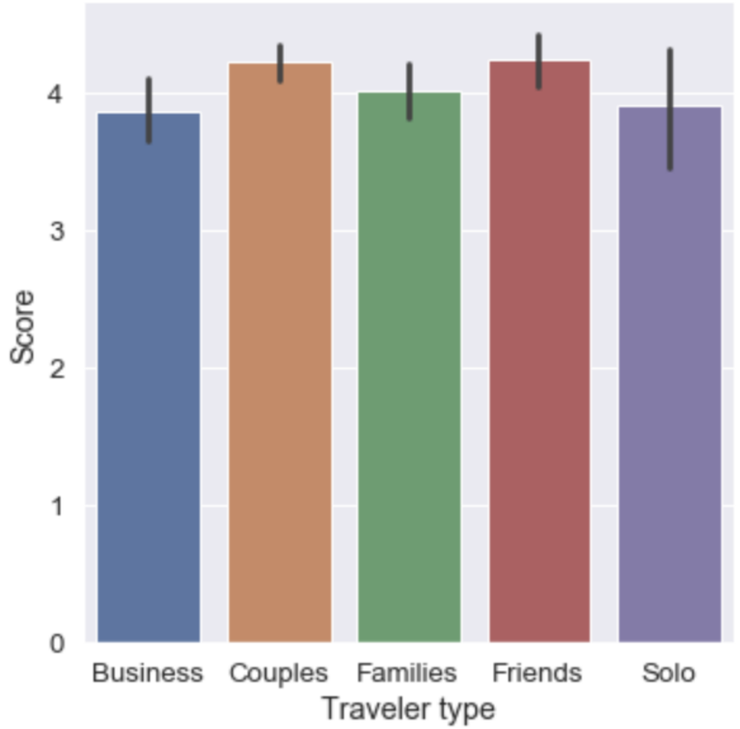
Ordering your categories

```
reviews["Traveler type"] = reviews["Traveler type"].astype("category")
reviews["Traveler type"].cat.categories
```

```
Index(['Business', 'Couples', 'Families', 'Friends', 'Solo'], dtype='object')
```

Updated visualization

```
sns.catplot(x="Traveler type", y="Score", data=reviews, kind="bar")
```



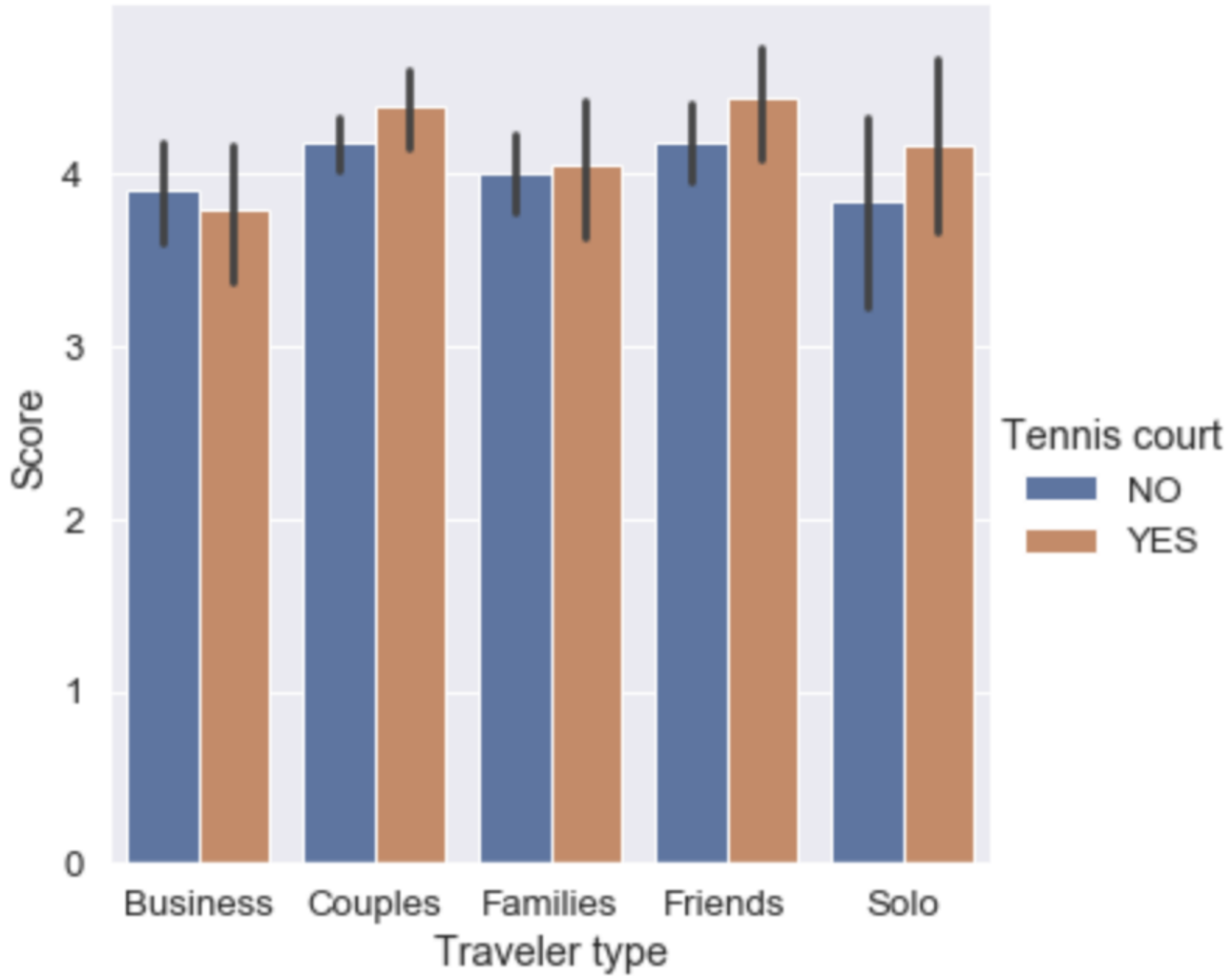
- Note: `catplot()` has an `order` parameter

The hue parameter

- `hue` :
  - name of a variable in `data`
  - used to split the data by a second category
  - also used to color the graphic

```
sns.set(font_scale=1.2)
sns.set_style("darkgrid")
sns.catplot(x="Traveler type", y="Score", data=reviews, kind="bar",
            hue="Tennis court")  # <--- new parameter
```

Bar plot across two variables



Bar plot practice

WORKING WITH CATEGORICAL DATA IN PYTHON

datacamp

Point and count plots

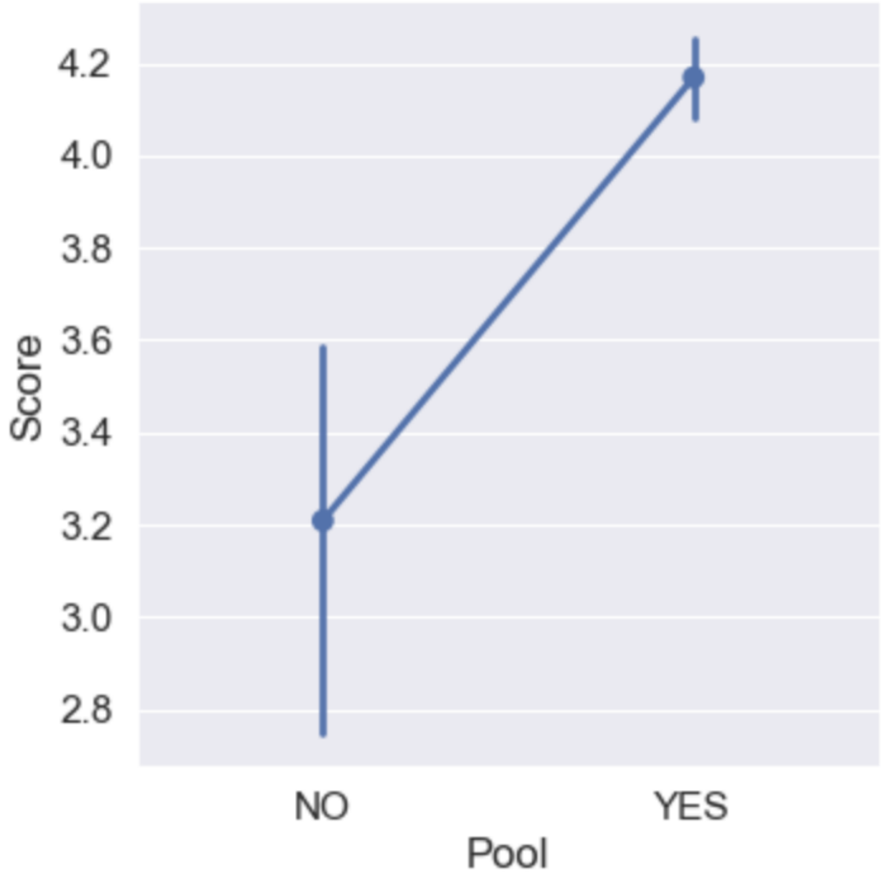
WORKING WITH CATEGORICAL DATA IN PYTHON

Kasey Jones
Research Data Scientist

datacamp

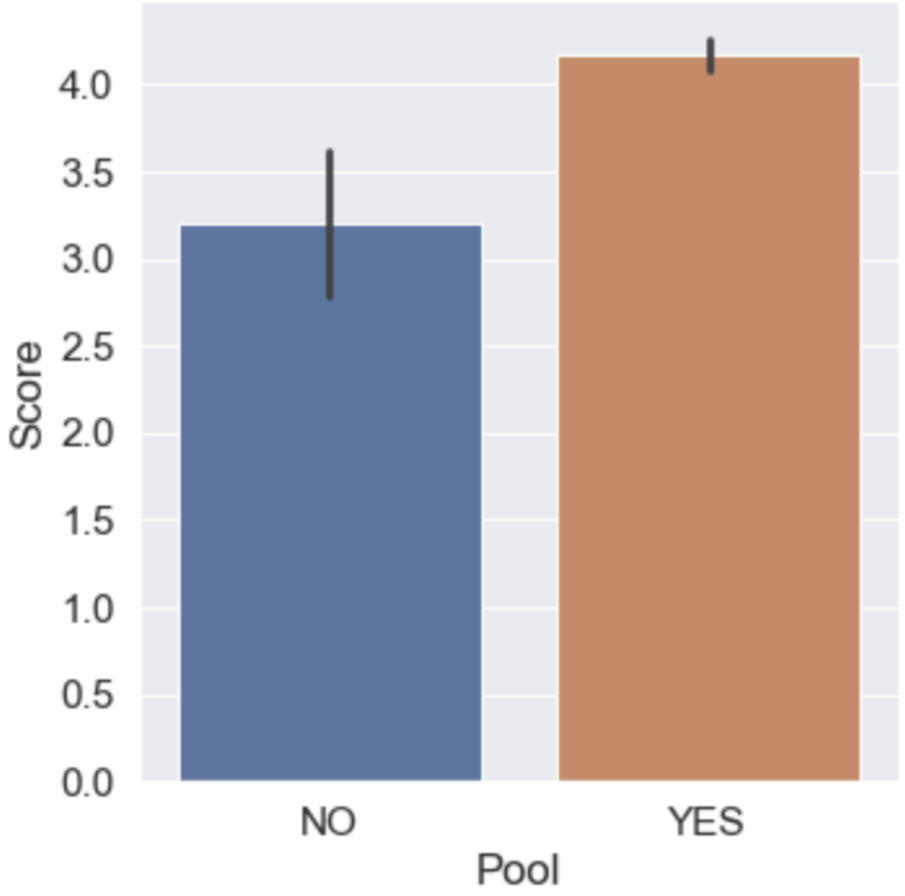
Point plot example

```
sns.catplot(x="Pool", y="Score", data=reviews, kind="point") # <--- updated
```

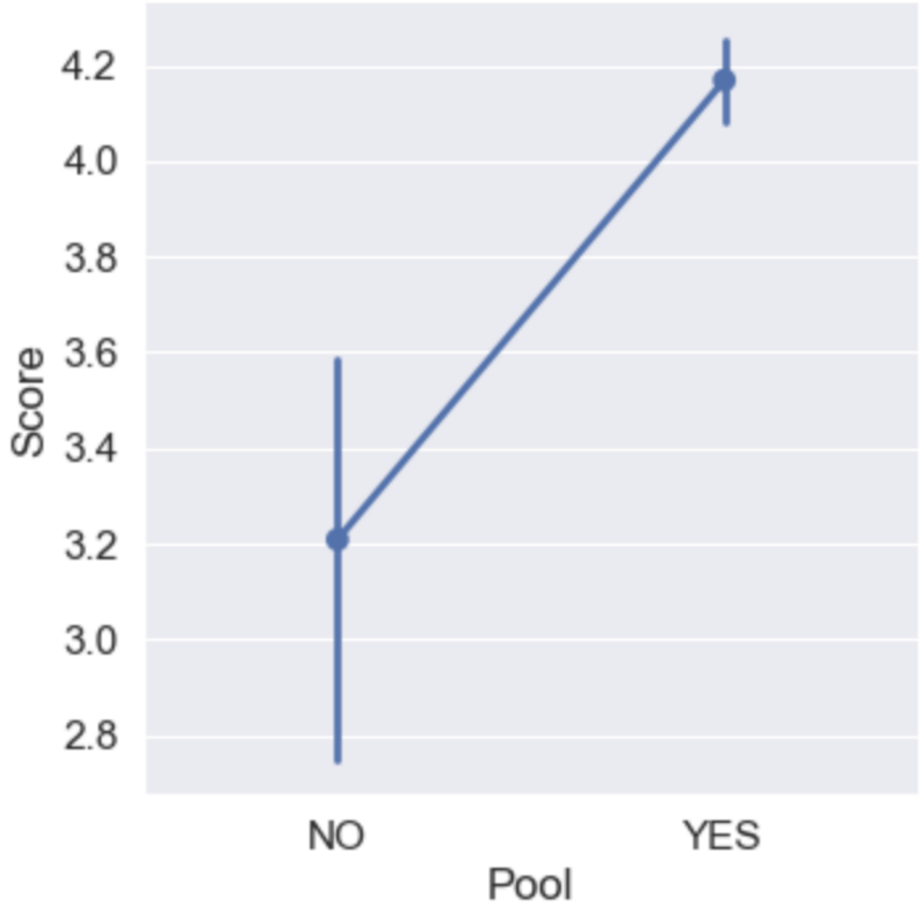


Bar plot vs. point plot

Bar plot

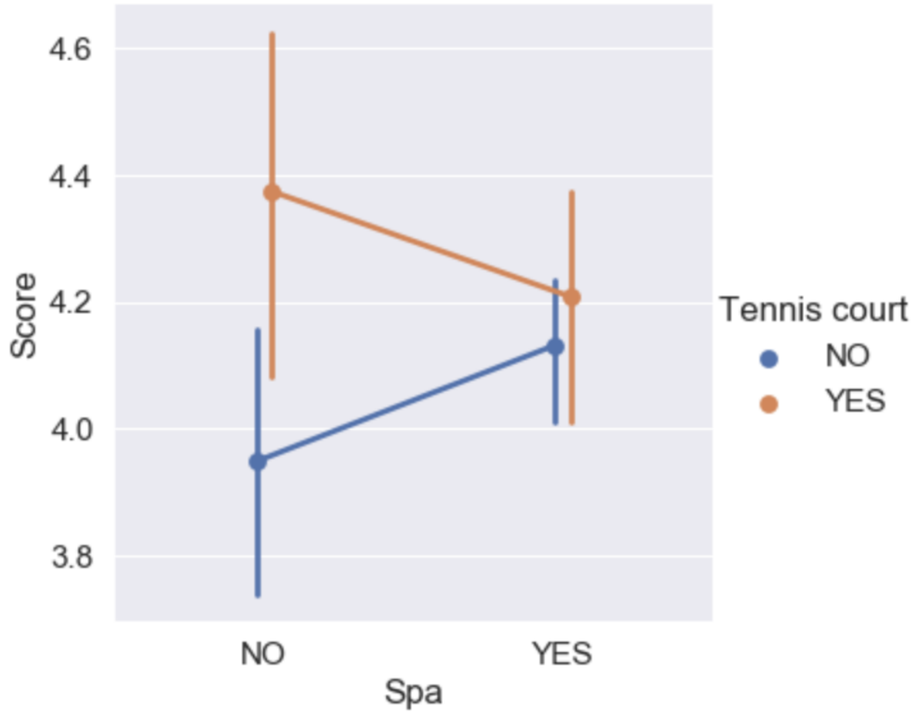


Point plot



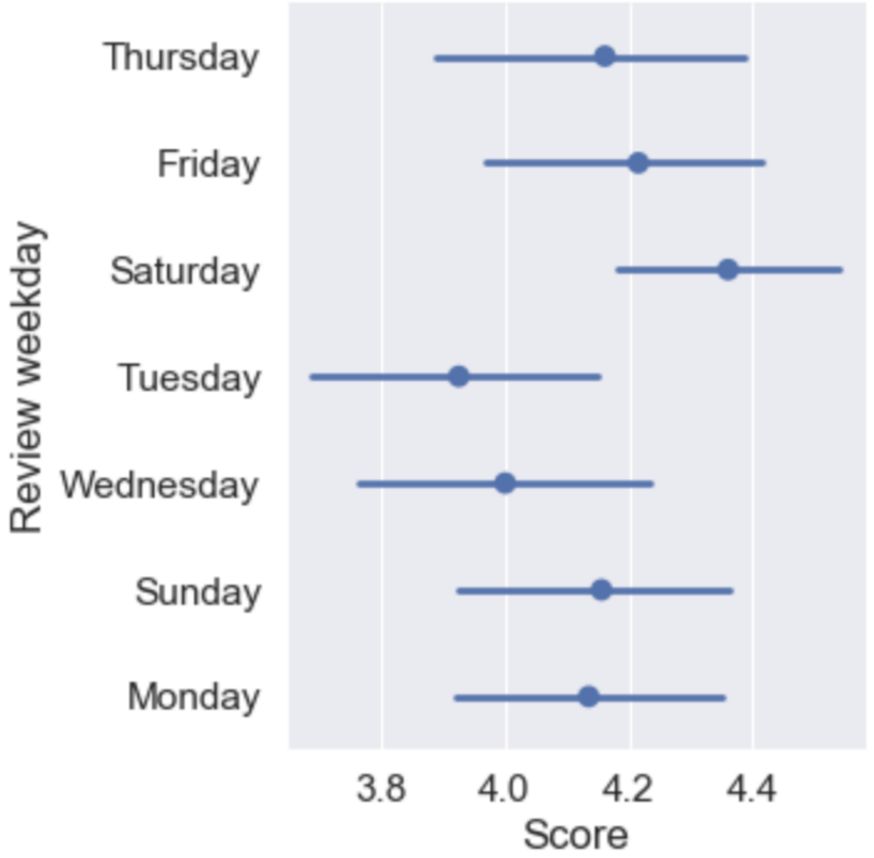
Point plot with hue

```
sns.catplot(x="Spa", y="Score", data=reviews, kind="point",
            hue="Tennis court", dodge=True  # < --- New Parameter!
)
```



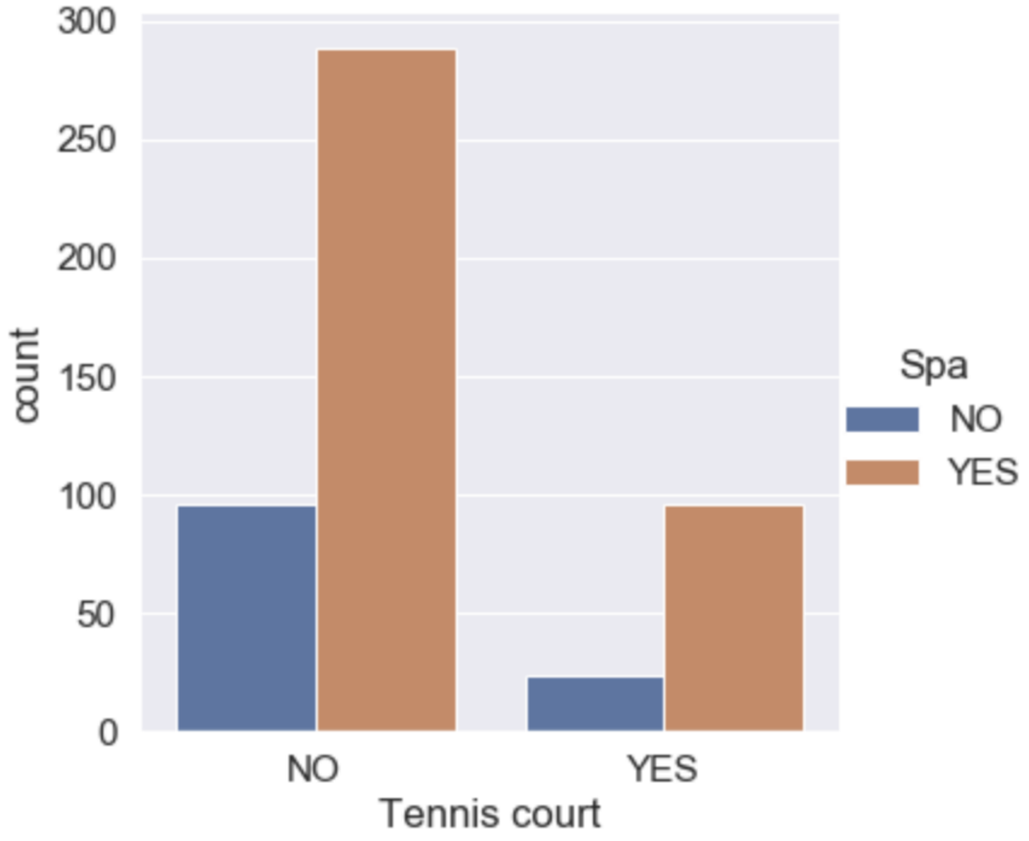
Using the join parameter

```
sns.catplot(x="Score",
            y="Review weekday",
            data=reviews,
            kind="point",
            join=False # < --- New!
)
```



One last catplot type

```
sns.catplot(x="Tennis court", data=reviews, kind="count", hue="Spa")
```



Time to practice!

WORKING WITH CATEGORICAL DATA IN PYTHON

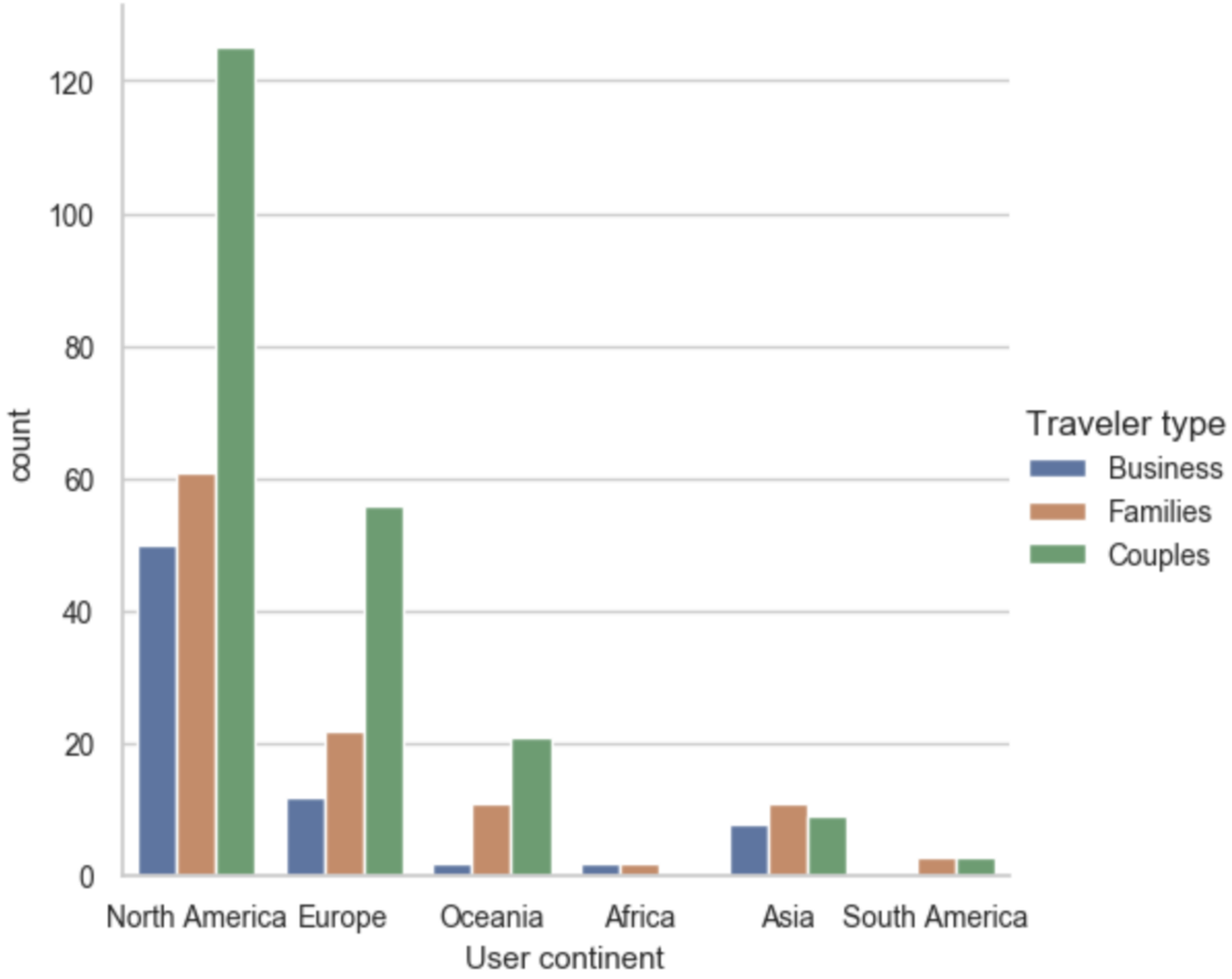
Additional catplot() options

WORKING WITH CATEGORICAL DATA IN PYTHON

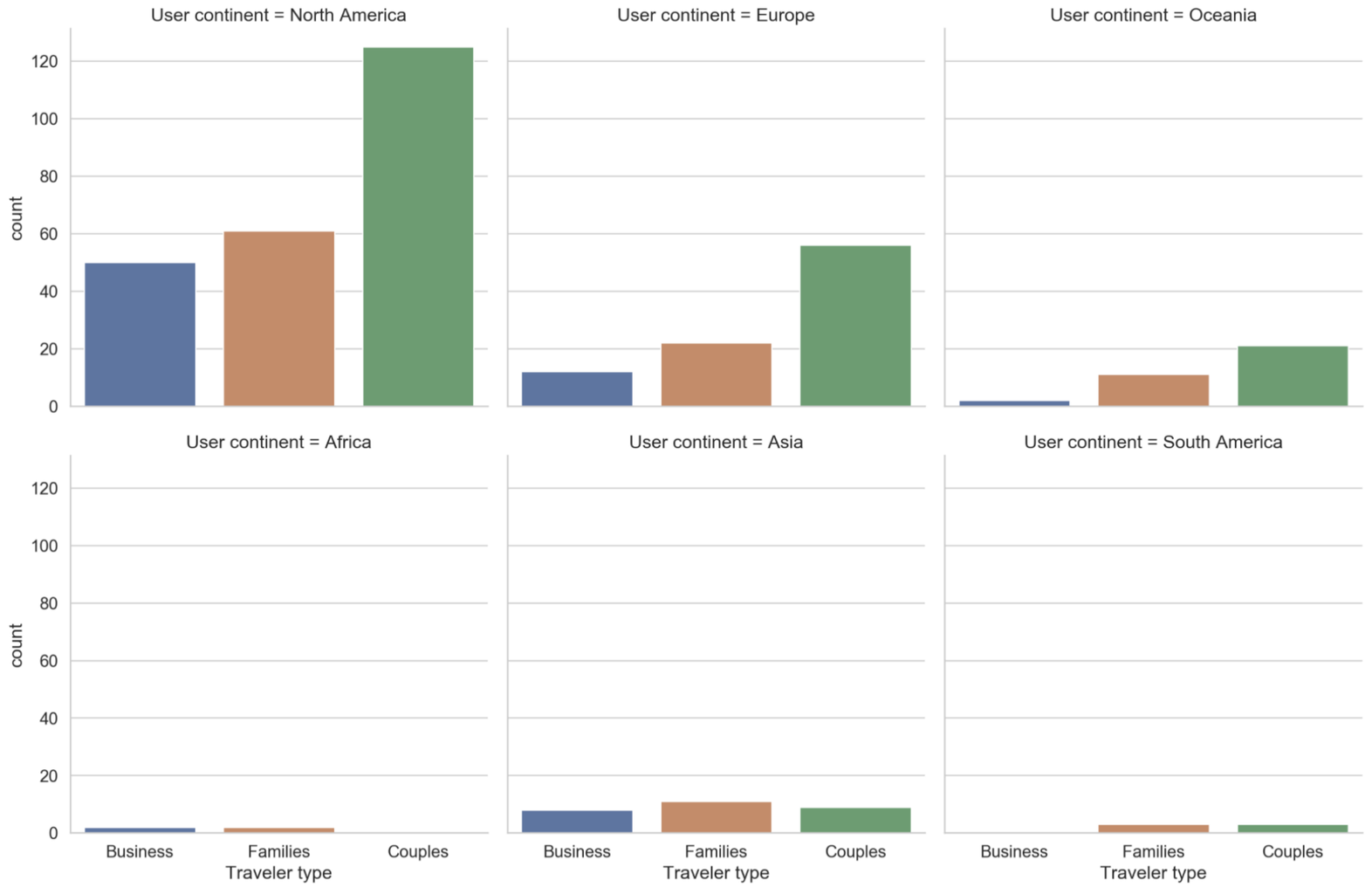
Kasey Jones
Research Data Scientist

datacamp

Difficulties with categorical plots



Using the catplot() facetgrid



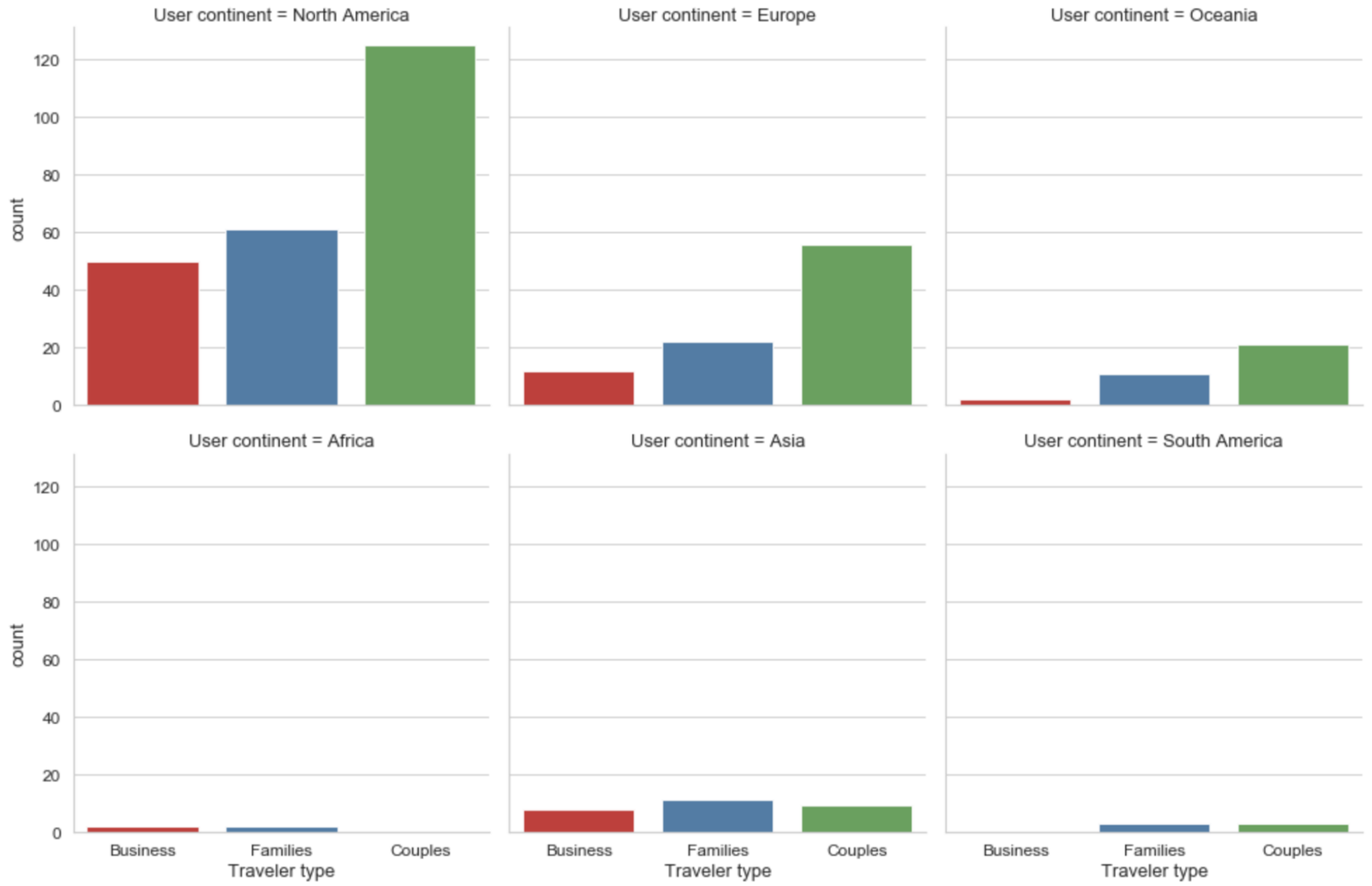
Using different arguments

```
sns.catplot(x="Traveler type", kind="count",
    col="User continent",
    col_wrap=3,
    palette=sns.color_palette("Set1"), data=reviews)
```

- `x` : `"Traveler type"`
- `kind` : `"count"`
- `col` : `"User continent"`
- `col_wrap` : `3`
- `palette` : `sns.color_palette("Set1")`
- Common colors: `"Set"` , `"Set2"` , `"Tab10"` , `"Paired"`

[1] http://seaborn.pydata.org/tutorial/color_palettes.html

One more look



Updating plots

- Setup: save your graphic as an object: `ax`
- Plot title: `ax.fig.suptitle("My title")`
- Axis labels: `ax.set_axis_labels("x-axis-label", "y-axis-label")`
- Title height: `plt.subplots_adjust(top=.9)`

```
ax = sns.catplot(x="Traveler type", col="User continent", col_wrap=3,
    kind="count", palette=sns.color_palette("Set1"), data=reviews)
ax.fig.suptitle("Hotel Score by Traveler Type & User Continent")
ax.set_axis_labels("Traveler Type", "Number of Reviews")
plt.subplots_adjust(top=.9)
plt.show()
```

Finished product

catplot() practice

WORKING WITH CATEGORICAL DATA IN PYTHON

datacamp