**Visualizing World Cup Data With Seaborn**

For this project you will be exploring data from the [Fifa World Cup](https://en.wikipedia.org/wiki/FIFA_World_Cup) from 1930-2014 to analyze trends and discover insights about the world's game, fútbol!

This Fifa World Cup data is from [Kaggle](https://www.kaggle.com/abecklas/fifa-world-cup). Kaggle is a platform for data science competitions that hosts many datasets online.

Using Seaborn you will create a series of plots that explore aggregates and distribution across the goals scored in World Cup games.

**A little primer on the Fifa World Cup:**

The FIFA World Cup, or simply the World Cup, is an international fútbol competition where 32 countries qualify to send teams made up of the best players from that nation to compete against each other for the World Cup championship.

The World Cup championship has been awarded every four years since the inaugural tournament in 1930, except in 1942 and 1946 when it was not held because of the Second World War.

The current format of the tournament involves 32 teams competing for the title at venues within the host nation over a period of one month.

**A note on datasets**

You may notice some typographical errors in the data as you view it. This is a large dataset, and errors are often a part of the process of generating large datasets. None of the errors should affect any of the steps of this project.

If you get stuck during this project, check out the **project walkthrough video** which can be found at the bottom of the page after the final step of the project.

### Visualize

1.

Import the modules that you'll be using in this project:

* import codecademylib3\_seaborn
* from matplotlib import pyplot as plt
* import pandas as pd
* import seaborn as sns

2.

Inspect the raw CSV files that you will be using in this project by selecting them in the file navigator.

3.

Load **WorldCupMatches.csv** into a DataFrame called df. This will allow you to eventually plot the DataFrame with Seaborn.

4.

It is usually a good idea to check any new DataFrame to make sure the results are as expected.

Inspect the DataFrame using .head(). Make sure to use print() to wrap any output you want to inspect.

5.

The data in **WorldCupMatches.csv** has the goals scored in each match broken up by goals for the home team and goals for the away team. We want to visualize the total number of goals scored in each match.

Create a new column in df named Total Goals, and set it equal to the sum of the columns Home Team Goals and Away Team Goals.

Print the results of df.head() to confirm your new column.

6.

You are going to create a bar chart visualizing how many goals were scored each year the World Cup was held between 1930-2014.

Set the style of your plot to be whitegrid . This will add gridlines to the plot which will make it easier to read the visualization.

7.

To make the text in your visualization bigger and easier to read, set the context to be poster.

If you would like to further adjust the font size of your plot, you can pass sns.set\_context() a second optional argument using the keyword font\_scale.

8.

Create a figure and axes for your plot using the syntax:

f, ax = plt.subplots()

Inside of plt.subplots(), set the size of the figure to be 12 inches wide and 7 inches tall.

9.

Using the data in df and the syntax:

ax = sns.barplot()

visualize the columns Year and Total Goals as a bar chart.

Year should be on the x-axis, and Total Goals should be on the y-axis.

10.

Render your bar chart so you can see it.

11.

Effective visualizations include a clear title.

Give your bar chart a meaningful title using ax.set\_title().

12.

Now you are going to create a box plot so you can visualize the distribution of the goals data instead of just the average with a bar chart.

Load **goals.csv** into a DataFrame called df\_goals, and take a quick look at the DataFrame using .head().

13.

Experimenting with different contexts and font scales can help you decide on the best context and font scale for the particular visualization.

Try setting the context of the plot to be notebook and the font\_scale to be 1.25.

14.

Create a figure for your second plot.

Set the variables f, ax2 and instantiate a figure that is 12 inches wide and 7 inches tall.

15.

Set ax2 equal to a box plot with the color palette Spectral that visualizes the data in the DataFrame df\_goals with the column year on the x-axis and goals on the y-axis.

16.

Give your box plot a meaningful and clear title.

17.

Render your box plot so you can see it.

18.

Congrats you're done! Feel free to continue iterating on your plots in this workspace.

You can also explore more datasets at the [Kaggle website](https://www.kaggle.com/).

19.

If you are stuck on the project or would like to see an experienced developer work through the project, watch the following project walkthrough video! <https://youtu.be/P4fxCeeKykQ>

import codecademylib3\_seaborn  
from matplotlib import pyplot as plt  
import pandas as pd  
import seaborn as sns  
  
df = pd.read\_csv('WorldCupMatches.csv')  
  
df["Total Goals"] = df["Home Team Goals"] + df["Away Team Goals"]  
  
print(df.head())  
  
# Bar Chart  
sns.set\_style("whitegrid")  
sns.set\_context("poster", font\_scale=0.5)  
  
f, ax = plt.subplots(figsize=(12,7))  
ax = sns.barplot(data=df, y="Total Goals", x="Year")  
ax.set\_title("Average Number of Goals Scored at Each World Cup")  
plt.show()  
  
  
df\_goals = pd.read\_csv('goals.csv')  
print(df\_goals.head())  
  
# Box Plot  
sns.set\_style("whitegrid")  
sns.set\_palette("Spectral")  
sns.set\_context("notebook", font\_scale=1.25)  
f, ax2 = plt.subplots(figsize=(12,7))  
ax2 = sns.boxplot(data=df\_goals, x='year', y='goals')  
ax.set\_title("Box Plot of Goals Scored at Each World Cup")  
plt.show()



