{"cells":[{"metadata":{},"cell_type":"markdown","source":"**This notebook is an exercise in the [Data Visualization](https://www.kaggle.com/learn/data-visualization) course. You can reference the tutorial at [this link](https://www.kaggle.com/alexisbcook/final-project).**\n\n---\n"},{"metadata":{},"cell_type":"markdown","source":"Now it's time for you to demonstrate your new skills with a project of your own!\n\nIn this exercise, you will work with a dataset of your choosing. Once you've selected a dataset, you'll design and create your own plot to tell interesting stories behind the data!\n\n## Setup\n\nRun the next cell to import and configure the Python libraries that you need to complete the exercise."},{"metadata":{"trusted":true},"cell_type":"code","source":"import pandas as pd\npd.plotting.register_matplotlib_converters()\nimport matplotlib.pyplot as plt\n%matplotlib inline\nimport seaborn as sns\nprint(\"Setup Complete\")","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"The questions below will give you feedback on your work. Run the following cell to set up the feedback system."},{"metadata":{"trusted":true},"cell_type":"code","source":"# Set up code checking\nfrom learntools.core import binder\nbinder.bind(globals())\nfrom learntools.data_viz_to_coder.ex7 import *\nprint(\"Setup Complete\")","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"## Step 1: Attach a dataset to the notebook\n\nBegin by selecting a CSV dataset from [Kaggle Datasets](https://www.kaggle.com/datasets). If you're unsure how to do this or would like to work with your own data, please revisit the instructions in the previous tutorial.\n\nOnce you have selected a dataset, click on the **[+ Add Data]** option in the top right corner. This will generate a pop-up window that you can use to search for your chosen dataset. \n\n![ex6_search_dataset](https://i.imgur.com/cIIWPUS.png)\n\nOnce you have found the dataset, click on the **[Add]** button to attach it to the notebook. You can check that it was successful by looking at the **Data** dropdown menu to the right of the notebook -- look for an **input** folder containing a subfolder that matches the name of the dataset.\n\n \n<img src=\"https://i.imgur.com/nMYc1Nu.png\" width=30%>
\n \n\nYou can click on the carat to the left of the name of the dataset to double-check that it contains a CSV file. For instance, the image below shows that the example dataset contains two CSV files: (1) **dc-wikia-data.csv**, and (2) **marvel-wikia-data.csv**.\n\n \n<img src=\"https://i.imgur.com/B4sJkVA.png\" width=30%>
\n \n\nOnce you've uploaded a dataset with a CSV file, run the code cell below **without changes** to receive credit for your work!"},{"metadata":{"trusted":true},"cell_type":"code","source":"# Check for a dataset with a CSV file\nstep_1.check()","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"## Step 2: Specify the filepath\n\nNow that the dataset is attached to the notebook, you can find its filepath. To do this, begin by clicking on the CSV file you'd like to use. This will open the CSV file in a tab below the notebook. You can find the filepath towards the top of this new tab. \n\n![ex6_filepath](https://i.imgur.com/fgXQV47.png)\n\nAfter you find the filepath corresponding to your dataset, fill it in as the value for `my_filepath` in the code cell below, and run the code cell to check that you've provided a valid filepath. For instance, in the case of this example dataset, we would set\n```\nmy_filepath = \"../input/fivethirtyeight-comic-characters-dataset/dc-wikia-data.csv\"\n``` \nNote that **you must enclose the filepath in quotation marks**; otherwise, the code will return an error.\n\nOnce you've entered the filepath, you can close the tab below the notebook by clicking on the **[X]** at the top of the tab."},{"metadata":{"trusted":true},"cell_type":"code","source":"# Fill in the line below: Specify the path of the CSV file to read\nmy_filepath = \"../input/fivethirtyeight-comic-characters-dataset/dc-wikia-data.csv\"\n# Check for a valid filepath to a CSV file in a dataset\nstep_2.check()","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"## Step 3: Load the data\n\nUse the next code cell to load your data file into `my_data`. Use the filepath that you specified in the previous step."},{"metadata":{"trusted":true},"cell_type":"code","source":"# Fill in the line below: Read the file into a variable my_data\nmy_data = pd.read_csv(my_filepath, low_memory=False)\n\n# Check that a dataset has been uploaded into my_data\nstep_3.check()","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"**_After the code cell above is marked correct_**, run the code cell below without changes to view the first five rows of the data."},{"metadata":{"trusted":true},"cell_type":"code","source":"# Print the first five rows of the data\nmy_data.head()","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"## Step 4: Visualize the data\n\nUse the next code cell to create a figure that tells a story behind your dataset. You can use any chart type (_line chart, bar chart, heatmap, etc_) of your choosing!"},{"metadata":{"trusted":true},"cell_type":"code","source":"# Create a

plot\nmy_data.sort_values(by='APPEARANCES', ascending=False, inplace=True)\nmy_clean_data = my_data[0:11][['ALIVE', 'APPEARANCES']]\n\n# Set the width and height of the figure\nplt.figure(figsize=(10,6))\n\n# Add title\nplt.title(\"Awards per Author\")\n\n# Bar chart showing average arrival delay for Spirit Airlines flights by month\nsns.barplot(x=my_clean_data['ALIVE'], y=my_clean_data['APPEARANCES'])\n\n# Add label for vertical axis\nplt.ylabel(\"N of Awards\")\nplt.xlabel(\"Author\")\nplt.xticks(rotation=\"90\")\n\n# Check that a figure appears below\nstep_4.check()\nplt.show()","execution_count":null,"outputs":[]},{"metadata":{},"cell_type":"markdown","source":"## Keep going\n\nLearn how to use your skills after completing the micro-course to create data visualizations in a **[final tutorial](https://www.kaggle.com/alexisbcook/creating-your-own-notebooks)**."},{"metadata":{},"cell_type":"markdown","source":"---\n\n\n\n\n*Have questions or comments? Visit the [Learn Discussion forum](https://www.kaggle.com/learn-forum/161291) to chat with other Learners.*"}],"metadata":{"kernelspec":{"language":"python","display_name":"Python 3","name":"python3"},"language_info":{"pygments_lexer":"ipython3","nbconvert_exporter":"python","version":"3.6.4","file_extension":".py","codemirror_mode":{"name":"ipython","version":3},"name":"python","mimetype":"text/x-python"}},"nbformat":4,"nbformat_minor":4}