Telling Stories

Project due
Thursday, July 22 @ 11:59PM

Telling Stories (TS)

Description:

In this project, you will incorporate a contextual layer for three refined data queries and visualizations that you initially created in Modules 4 & 5 from the CSV file compiled by FiveThirtyEight (538) about superheroes. They wrote an article about this particular dataset here.

To complete this project, you'll need to integrate all of the skills you've learned this semester: Python basic and complex commands, e.g., variables, input, print statements, conditions, and boolean operators, functions, lists, loops and parameters, Pandas/Numpy data analysis, and plotting techniques using Matplotlib.

New to this project is the use of Markdown language to add text into Jupyter Notebook to describe and analyze your work produced in the Pandas, Numpy and Matplotlib libraries.

from query to plot to story

1. Ensure the data is correct. Adjust the data queries, if needed.

Review the data for any inconsistencies. Clean it by filtering out any ND (“no data” cells) or flagging that you will need to add some explanation about “dirty data” in your final analysis. Take a step back and consider whether your initial data queries are appropriate for the larger story you want to tell. This might also require returning to step 1 above to ensure the data for the new query is consistent.

2. Plot the data queries.

Illustrate the data with an appropriate graph, chart or formatted crosstab. While the plot might look good, it may not actually make sense to the reader or it may not help advance the story. Cast a critical eye on this step of the process and be willing to try another view.

3. Determine the data context.

Describe the story by providing context and background about the data and plots. Explain why your findings matter. Is there an historical or cultural explanation for the results? Summarize the computational process that led to the final data analysis, and your conclusions.

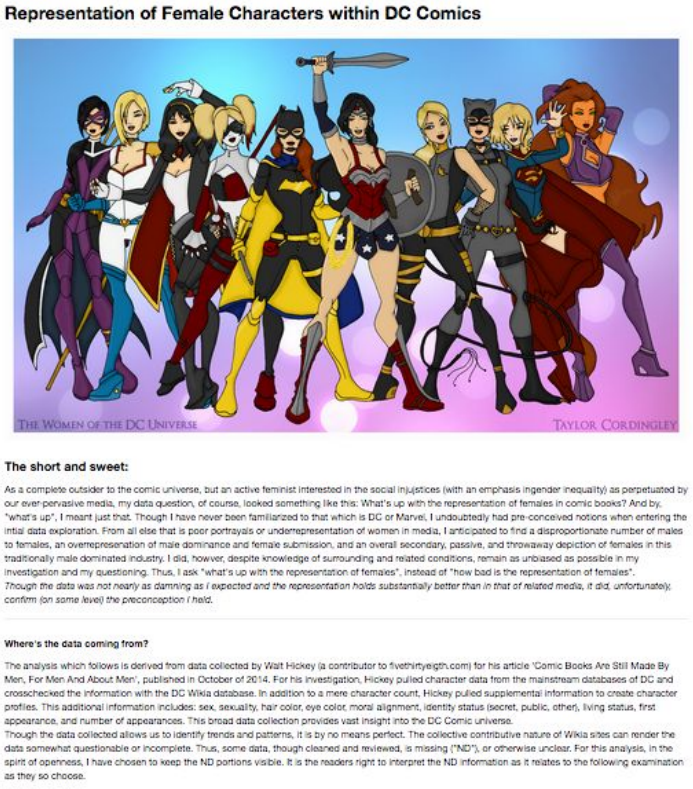
writing the data narrative, part 1

Brief description of your research query

(2-3 paragraphs)

Describe the overall question or theme of your analysis.

Describe the CSV file that you analyzed. What is it? Where did it come from? How was the data collected? What data is included? What data is missing?



Representation of Female Characters within DC Comics

The short and sweet:

As a complete outsider to the comic universe, but an active feminist interested in the social injustices (with an emphasis ingender inequality) as perpetuated by our ever-pervasive media, my data question, of course, looked something like this: What's up with the representation of females in comic books? And by, "what's up", I meant just that. Though I have never been familiarized to that which is DC or Marvel, I undoubtedly had pre-conceived notions when entering the initial data exploration. From all else that is poor portrayals or underrepresentation of women in media, I anticipated to find a disproportionate number of males to females, an overrepresenation of male dominance and female submission, and an overall secondary, passive, and throwaway depiction of females in this traditionally male dominated industry. I did, however, despite knowledge of surrounding and related conditions, remain as unbiased as possible in my investigation and my questioning. Thus, I ask "what's up with the representation of females", instead of "how bad is the representation of females". *Though the data was not nearly as damning as I expected and the representation holds substantially better than in that of related media, it did, unfortunately, confirm (on some level) the preconception I held.*

Where's the data coming from?

The analysis which follows is derived from data collected by Walt Hickey (a contributor to fivethirtyeigth.com) for his article 'Comic Books Are Still Made By Men, For Men And About Men', published in October of 2014. For his investigation, Hickey pulled character data from the mainstream databases of DC and crosschecked the information with the DC Wikia database. In addition to a mere character count, Hickey pulled supplemental information to create character profiles. This additional information includes: sex, sexuality, hair color, eye color, moral alignment, identity status (secret, public, other), living status, first appearance, and number of appearances. This broad data collection provides vast insight into the DC Comic universe.
Though the data collected allows us to identify trends and patterns, it is by no means perfect. The collective contributive nature of Wikia sites can render the data somewhat questionable or incomplete. Thus, some data, though cleaned and reviewed, is missing ("ND"), or otherwise unclear. For this analysis, in the spirit of openness, I have chosen to keep the ND portions visible. It is the readers right to interpret the ND information as it relates to the following examination as they so choose.

writing the data narrative, part 2

Explain your CT approach to creating the queries and plots
(2-3 paragraphs)

Describe the different computational principles (expression, etc.) and practices (abstraction, etc.) that you used to conduct the data analysis.

Where's my mind?

It is important to disclose the practices and perspectives utilized to conduct this data dive. Obviously, there is the basics. This, being my first crack at DIY data visuals, was wrought with testing and debugging. At every step I was runing the imports, filtering the data in new ways, creating baby data sets, dropping unnecessary clutter, and trying to keep things pretty. When I ran into catastrophic error messages, I would have to review, research, adjust and try again. In this experimentation, though, lies the fun. This project uses remixed and reused code from many helpful resources (shoutouts at the end) in order to create something new and specific. Borrowing the data from Hickey, I was able to create my own analysis and paint my own picture. Abstracting and modularizing allowed me to dive into one topic, and relate it back to a bigger picture to determine what information was meaningful, and what wasnt. It allowed me to ask three separate questions to complete one overarching overview.
And as always, this project could not have succeeded without the questioning mentality. As mentioned before, it is impossible to not arrive with bias as you are forever connected to that which you already "know" and experience. The questioning mindset, though, allows you to move beyond the connection and preconceived notions, to truly explore something new. In this piece, I was able to use something so foreign, to dive into a topic unknown (by me), and create an expressive, and hopefully unique data analysis.

writing the data narrative, part 3

Identify the data question

(2-3 paragraphs for **each** computation)

Use Markdown language and formatting to clearly describe each data question and the computational approach you used to answer them.

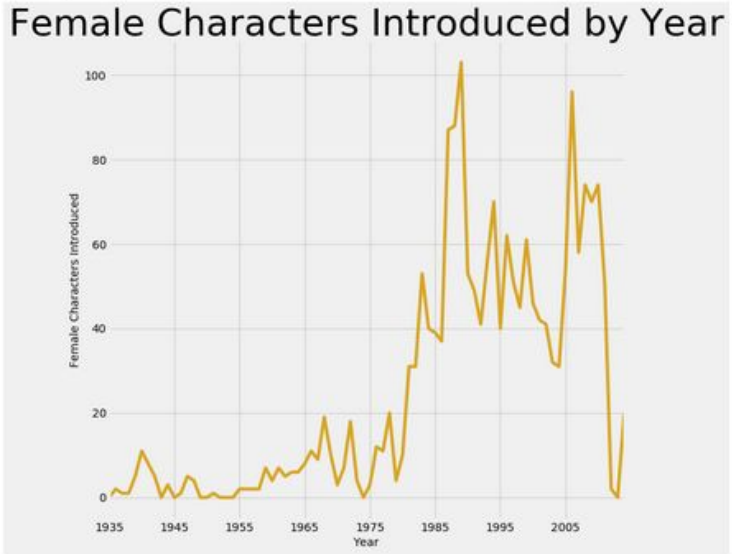
If you opted to leave the ND and/or low number of instances in your analysis, explain why and what effect you think it may have on the findings. If you dropped some data, be sure to acknowledge that in the story too.

FEMALE CHARACTERS INTRODUCED BY YEAR

1. How many female characters were released in every given year of DC Comics?

In addition to representation information, I wanted to evaluate the introduction years of the female DC characters. To answer questions on patterns and trends, as well as assessing the social circumstances which may prompt a peak in female character development, I needed to see a data visualization which would make it easy to identify peaks and troughs. Unlike the previous questions, this question doesnt need a comparative analysis, for the information is meaningful on its own. Our anaysis lies in the research or knowledge of the womens movement or other likely causes of the years which saw the greatest numbers of introductions.
The best way to produce an easy-to-read timeline was a line graph. The line graph below shows the total number of new female characters introduced per year from 1935 to 2010. From the graph we can easily identify the highs and lows.

In [11]:
```
#FEMALES BY YEAR
yearly_female = pd.crosstab(DCCOMICS["YEAR"], DCCOMICS["SEX"] == "Female Characters")
yearly_femaleonly = yearly_female[[1]].rename(columns={True: "Female"})
#PLOT OF FEMALES BY YEAR
yearly_femaleonly.plot(legend = None, color="goldenrod")
plt.title("Female Characters Introduced by Year", fontsize = 45)
plt.ylabel("Female Characters Introduced", fontsize = 14)
plt.xlabel("Year", fontsize = 14)
plt.title("Female Characters Introduced by Year", fontsize = 45, ha = "center")
plt.show()
```



From the onset of comic book creation, the implementation of female characters remains extremely low. Few (and even none in some years) female characters were released, or utilized in comic book series'. Beginning in 1967 we begin to see more movement, and in 1979 begins a major upward trend. Female character introduction reached its peak in 1989, and sees an additional peak in 2006. From there, we saw a drastic drop leading to the end of this data.

writing the data narrative, part 4

Summarize your findings and any limitations of the data
(2-3 paragraphs)

Describe the overall results of your analysis. Provide a brief synopsis of the data questions and the results?

Include any barriers to your discovery, e.g., dirty data, lack of data, in-depth questions vs novice skill level, etc.

What do you think?

As I mentioned before, this information was far less bleak than I anticipated. Though females are DRASTICALLY underrepresented (we do make up 51% of the population, afterall), and even more underrepresented when you remove the female one hit wonders, their representation aligns closely with male characters. We dont see a particularly high moral ambiguity (just a bit higher than their counterparts) or any outlandish trends among identity status. In fact, the female characters were MORE likely to be public with their heroism! While the figures remain discouraging and disproportionate, we can at least take solace that the 29% of characters which are female, are experience relative parity in representation.
Moving forward, I would hope to see the gender alignment of DC Comics characters more closely resemble the real world. Shoutout to the one transgender character! Now lets see some more!

Food for thought:

The following, is one of my all-time favorite TED Talks. Christopher Bell speaks on the current (2015) configuration of the superhero industry. But his issue, to my surprise, was not with the number or representation, so much as the associated merchandise and toys. Please enjoy!

```
In [12]: from IPython.display import HTML
         HTML('<iframe width="560" height="315" src="https://www.youtube.com/embed/0_z2v427G20" frameborder="0" allowfullscreen>
```

Out[12]:

writing the data narrative, part 5

acknowledgments
(1 paragraph)

Note any sources of help, e.g., websites, books, forums, peers, etc.

Acknowledgements

So many people to thank for help!
First, a huge thank you to Walt Hickey for inspiration and data! His article can be found here: https://fivethirtyeight.com/features/women-in-comic-books/
A shoutout to Wendy, Steven and the INFO1201 class for all of their help.
Thanks to: http://pandas.pydata.org/pandasdocs/version/0.15.0/visualization.html#visualization-barplot for showing me my options!
And to: http://www.futurile.net/2016/02/27/matplotlib-beautiful-plots-with-style/ for giving me the awesome fivethirtyeight.com formatting style for my graphs!
Thanks to: https://matplotlib.org/examples/color/named_colors.html for all the color inspo!
Thanks to Taylor Cordingley: http://femmes-fatales.deviantart.com/art/The-Women-of-the-DC-Universe-Redesigned-518910256,
DCComics:http://www.dccomics.com/characters/wonder-woman, and Adam Hughes throback poster: http://femmes-fatales.deviantart.com/art/The-Women-of-the-DC-Universe-Redesigned-518910256 for the artwork!
Thanks to the DC Comic creators who recognize the power of female superhero's!
AND THANKS TO THE FEMALE SUPERHEROES WHO KICK SO MUCH ASS!!

project Rubric

Submission:

1. Make sure that your project works.
   Try running it a few times.
   Have a friend play with it to see if it makes sense to them, re: Computational perspectives: Expression, connection, intention, etc.
2. Double check the project rubric to make sure you've fulfilled all the requirements.
3. Submit 2 files to Canvas:

   LastName.TS.ipynb and .html

4. Complete the project reflection form in the Canvas/Assignments folder:

   Week 7

Rubric: 10 points total

3 pts: Plots **<u>at least 3 different</u>** data questions from the Mod 4 & 5 projects.

3 pts: Plot code is integrated with a narrative that describes the project. The Markdown language text should include:

- brief description of the overall research query/theme
- identify the data questions
- explain the computational process and results for **<u>each plot</u>**
- summarize the overall findings
- limitations about data that could call results into question
- acknowledges sources of help

3 pts: Code incorporates both simple and complex Python commands. Project uses the Pandas, Numpy, and Matplotlib libraries to perform the data analysis and plot the data.

1 pt: Code is organized and easy to read.

Extra Credit: 11 pts: For the select few projects that went above and beyond.