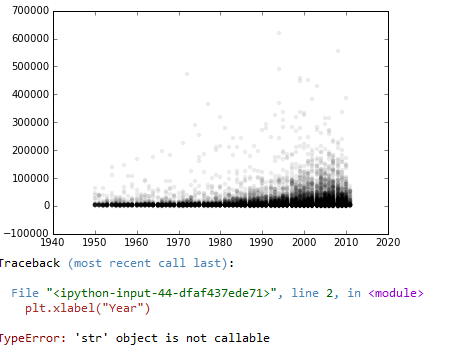
Pandas Homework (Class 5)

1. Looking at at vote count by year, vote count increases over time, but seems it starts to decrease in the most recent years. What am I doing wrong that my labels are not working?

plt.scatter(data.year, data.votes, lw=0, alpha=.08, color='k')

plt.xlabel("Year")

plt.ylabel("Vote Count")



1. Calculate counts and means for each genre (Not sure how to summarize and save into a table I can select from)

data.groupby('genres').genres.count()

Out[72]:

genres

Action 24

Action|Adventure 21

Action|Adventure|Animation|Comedy|Family|Fantasy 1

Action|Adventure|Animation|Family|Fantasy|Sci-Fi 1

Action|Adventure|Biography|Crime|History|Western 1

Action|Adventure|Biography|Drama|History 4

Action|Adventure|Biography|Drama|History|Romance|War 1

Action|Adventure|Biography|Drama|Romance|War 1

Action|Adventure|Biography|Drama|War 1

Action|Adventure|Biography|History|Musical|Romance 1

Action|Adventure|Biography|History|Romance|War 1

Action|Adventure|Biography|Romance|Western 1

Action|Adventure|Comedy 22

Action|Adventure|Comedy|Crime 9

Action|Adventure|Comedy|Crime|Drama 1

...

War|Drama|Action 1

War|Drama|Action|Romance 1

War|Drama|Romance 1

War|Thriller 1

War|Western 1

Western 24

Western|Action 1

Western|Action|Adventure 1

Western|Adventure|Drama 3

Western|Comedy 2

Western|Crime 1

Western|Drama 1

Western|Mystery|Crime|Action 1

Western|Romance|Drama 1

Western|Sci-Fi|Thriller 1

Name: genres, Length: 1643, dtype: int64

1. In comparing Action vs. Westerns, Westerns have higher average ratings and lower variation (std deviation).

data[data.genres=='Action'].score.describe()

Out[73]:

count 24.000000

mean 5.837500

std 1.260715

min 2.400000

25% 5.650000

50% 6.200000

75% 6.800000

max 7.100000

Name: score, dtype: float64

data[data.genres=='Western'].score.describe()

Out[74]:

count 24.000000

mean 7.041667

std 0.938045

min 4.800000

25% 6.550000

50% 7.200000

75% 7.500000

max 9.000000

1. Calculate mean for each genre

data.groupby('genres').genres.mean() #can’t see what I’m doing wrong here

Traceback (most recent call last):

File "<ipython-input-77-a4170dba06da>", line 1, in <module>

data.groupby('genres').genres.mean()

File "C:\Users\labraham\AppData\Local\Continuum\Anaconda\lib\site-packages\pandas\core\groupby.py", line 691, in mean

return self.\_cython\_agg\_general('mean')

File "C:\Users\labraham\AppData\Local\Continuum\Anaconda\lib\site-packages\pandas\core\groupby.py", line 1097, in \_cython\_agg\_general

raise DataError('No numeric types to aggregate')

DataError: No numeric types to aggregate