Histograms And Box Plots

Syntax

- Creating a frequency distribution:

```
norm_reviews['Fandango_RatingValue'].value_counts()
```

- Creating a histogram:

```
ax.hist(norm_reviews['Fandango_RatingValue'])
```

- Specifing the lower and upper range of bins within a histogram:

```
ax.hist(norm_reviews['Fandango_RatingValue'], range=(0,5))
```

- Setting y-axis limits:

```
ax.set_ylim(0,50)
```

- Setting number of bins for a histogram:

```
ax.hist(norm_reviews['Fandango_RatingValue'], bins = 20)
```

- Creating a box plot:

```
ax.boxplot(norm_reviews["RT_user_norm"])
```

- Creating a boxplot for multiple columns of data:

```
num_cols = ['RT_user_norm', 'Metacritic_user_nom', 'IMDB_norm', 'Fandango_Ratingvalue']

ax.boxplot(norm_reviews[num_cols].values)
```

Concepts

- Frequency distribution consists of unique values and corresponding frequencies.

- Bins are intervals of fixed length to cover all possible values.

- Histogram shows the distribution over numerical data.

- Quartiles divide the range of numerical values into four different regions.

- Box plot visually shows quartiles of a set of data as well as any outliers.

- Outliers are abnormal values that affect the overall observation of the data set due to their very high or low values.

Resources

- [Documentation for histogram](#)
- [Documentation for boxplot](#)
- [Various ways to show dsistributions](#)