MoneyBall

—

Nick Pipal

Introduction to the DataSet

- Where did I find this dataset?
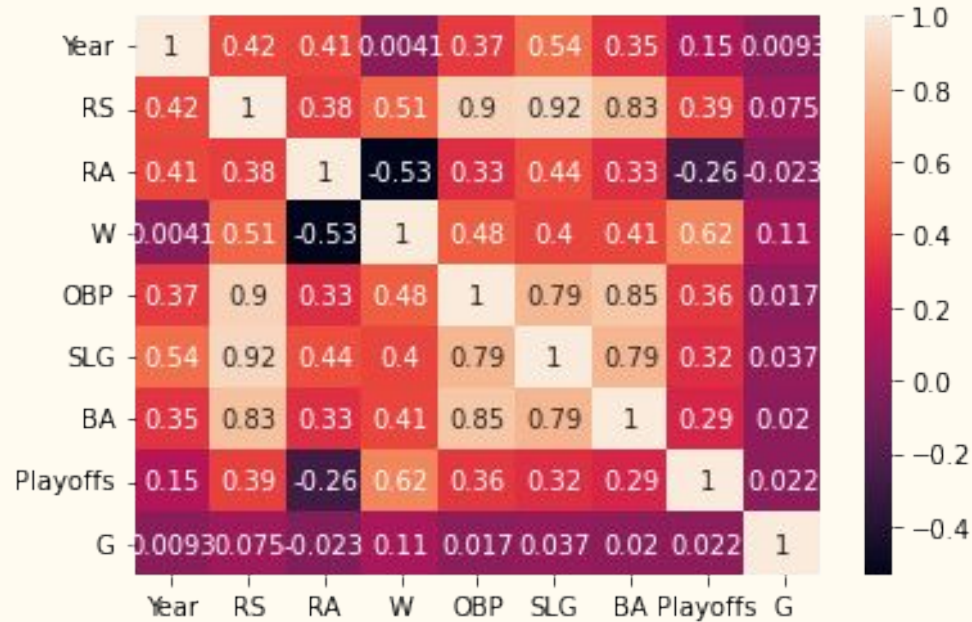

- What is the dataset about?
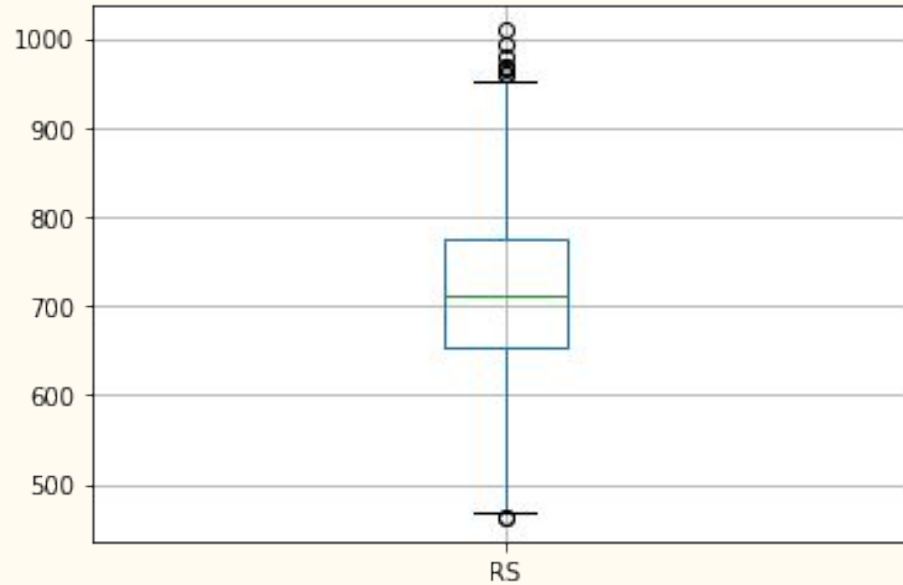
Inspecting/Cleaning/Removing Data

- Number of Columns
- Number of Rows
- Data Types
- Any duplicate values?
- Missing data?
- Any inconsistent values in Numerical or Object Columns?

```
RangeIndex: 1232 entries, 0 to 1231
Data columns (total 15 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   Team           1232 non-null    object
 1   League         1232 non-null    object
 2   Year           1232 non-null    int64
 3   RS             1232 non-null    int64
 4   RA             1232 non-null    int64
 5   W              1232 non-null    int64
 6   OBP            1232 non-null    float64
 7   SLG            1232 non-null    float64
 8   BA             1232 non-null    float64
 9   Playoffs       1232 non-null    int64
 10  RankSeason     244 non-null     float64
 11  RankPlayoffs   244 non-null     float64
 12  G              1232 non-null    int64
 13  OOBP           420 non-null     float64
 14  OSLG           420 non-null     float64
dtypes: float64(7), int64(6), object(2)
```
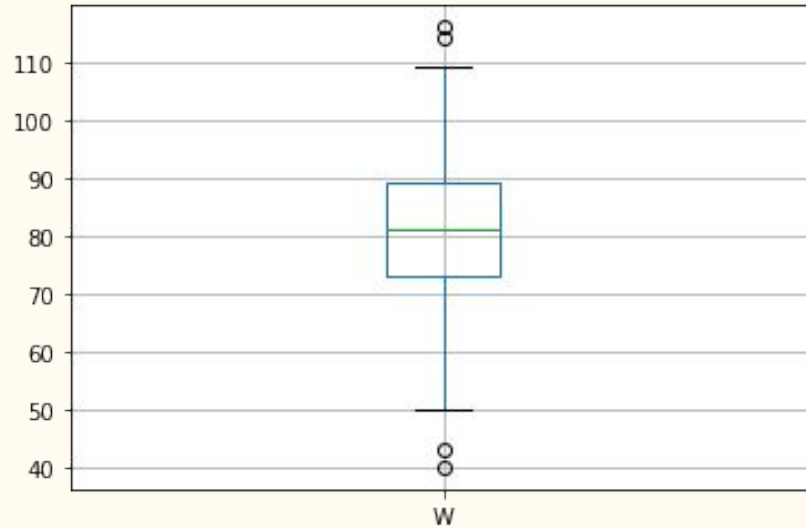
How can we better understand the data?
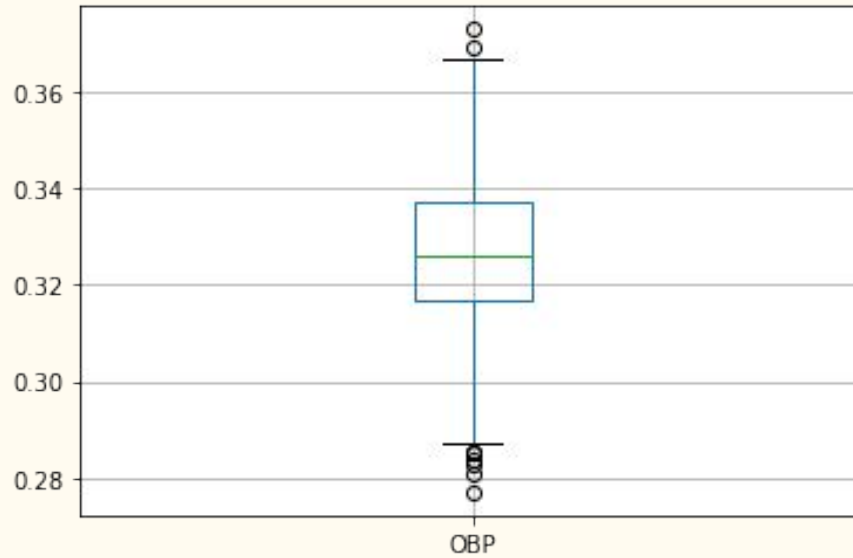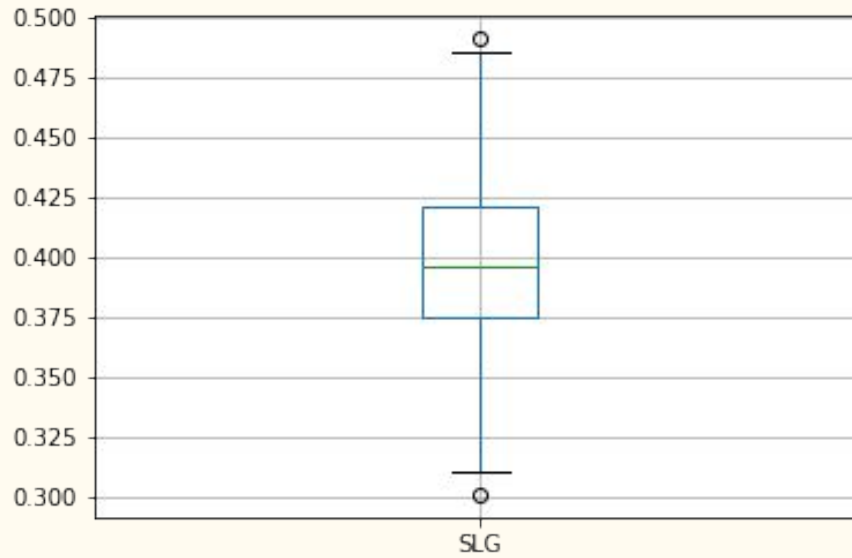
Exploratory Visuals

- A couple of outliers in the data for Runs Scored but everything is plausible
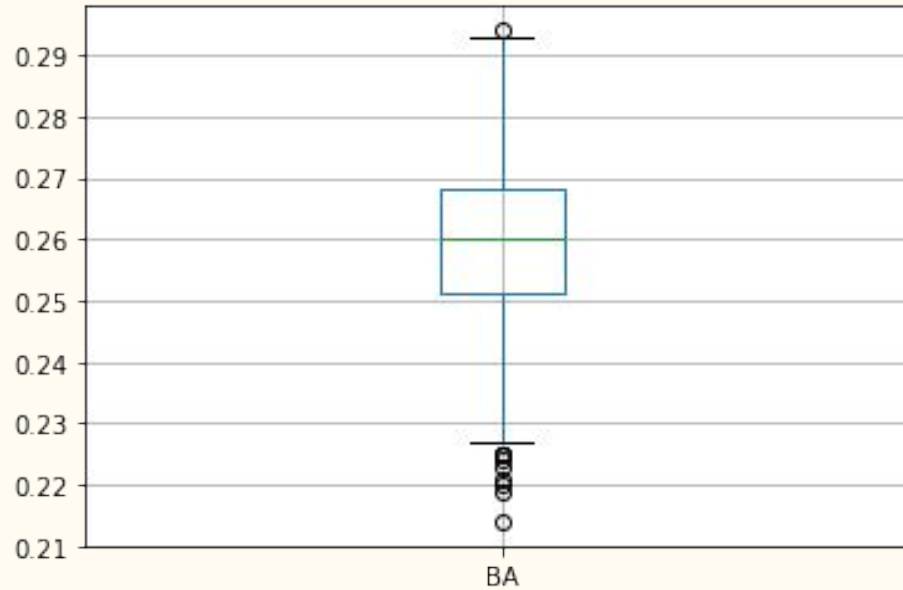
- A couple of outliers in the data for Wins but everything is plausible

- Some outliers in the data for On-Base-Percentage but everything is plausible

- A couple of outliers in the data for Slugging Percentage but everything is plausible

- Some outliers in the data for Batting Average but everything is plausible