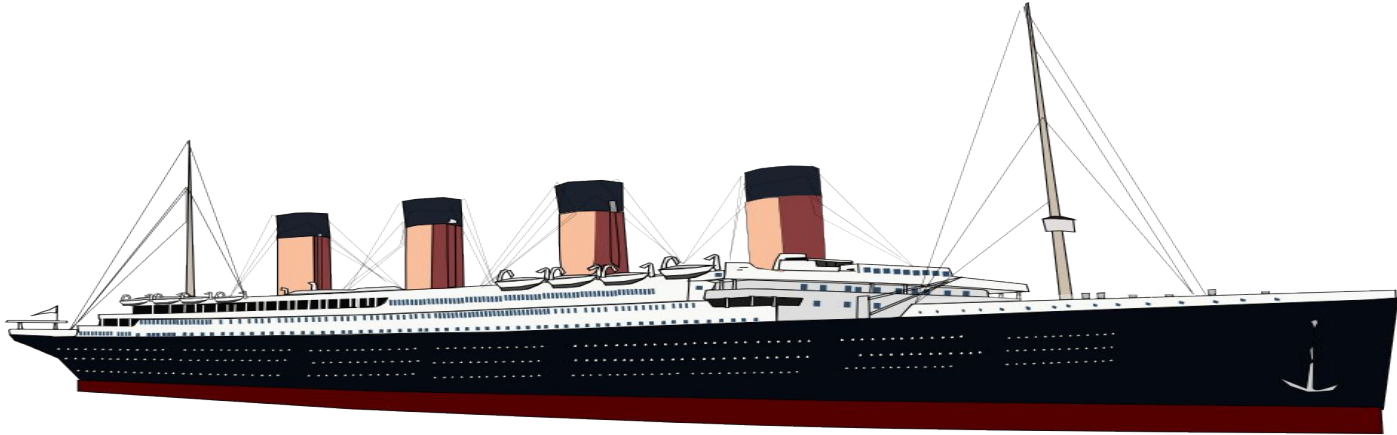


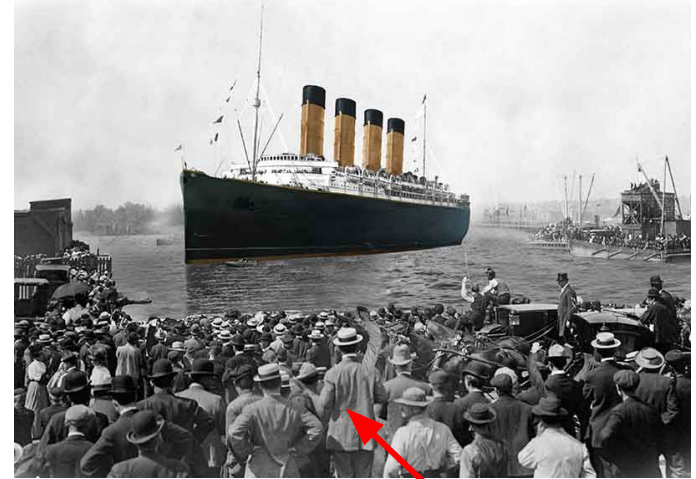
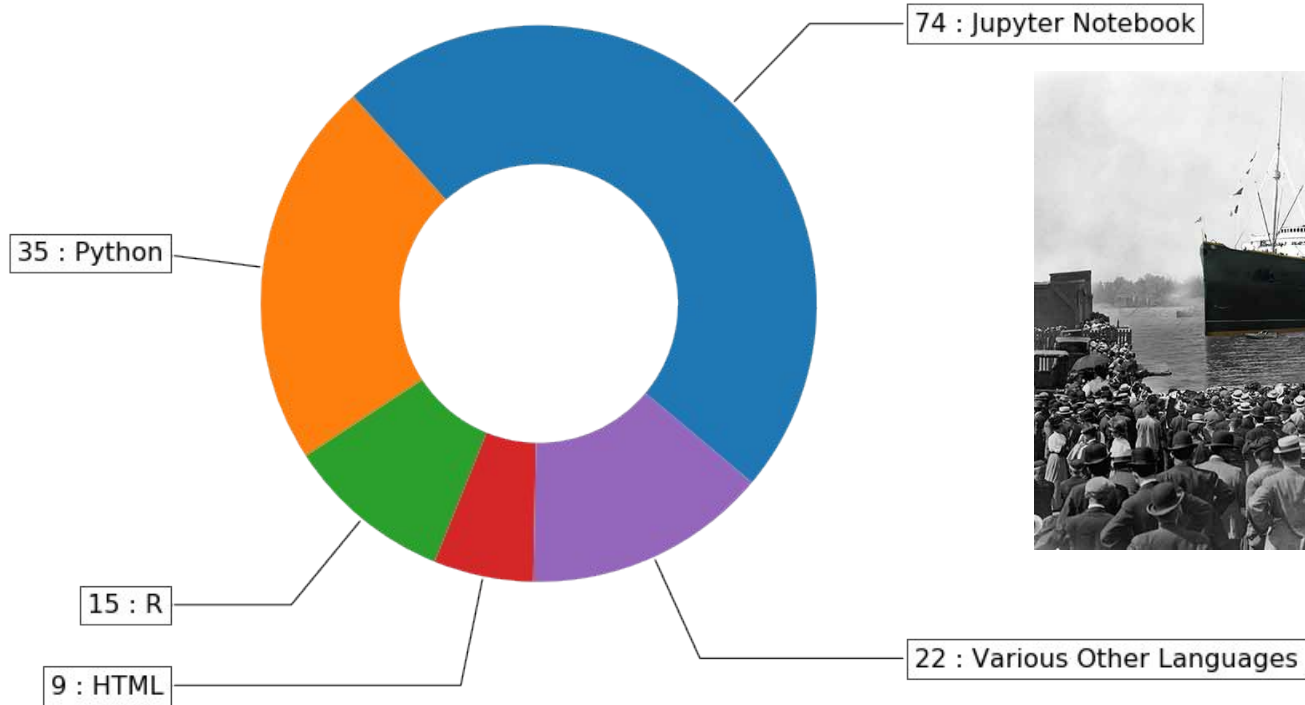
Can you predict the language of a repository using popular Titanic data sets?

Sara Peña & Paddy Doran



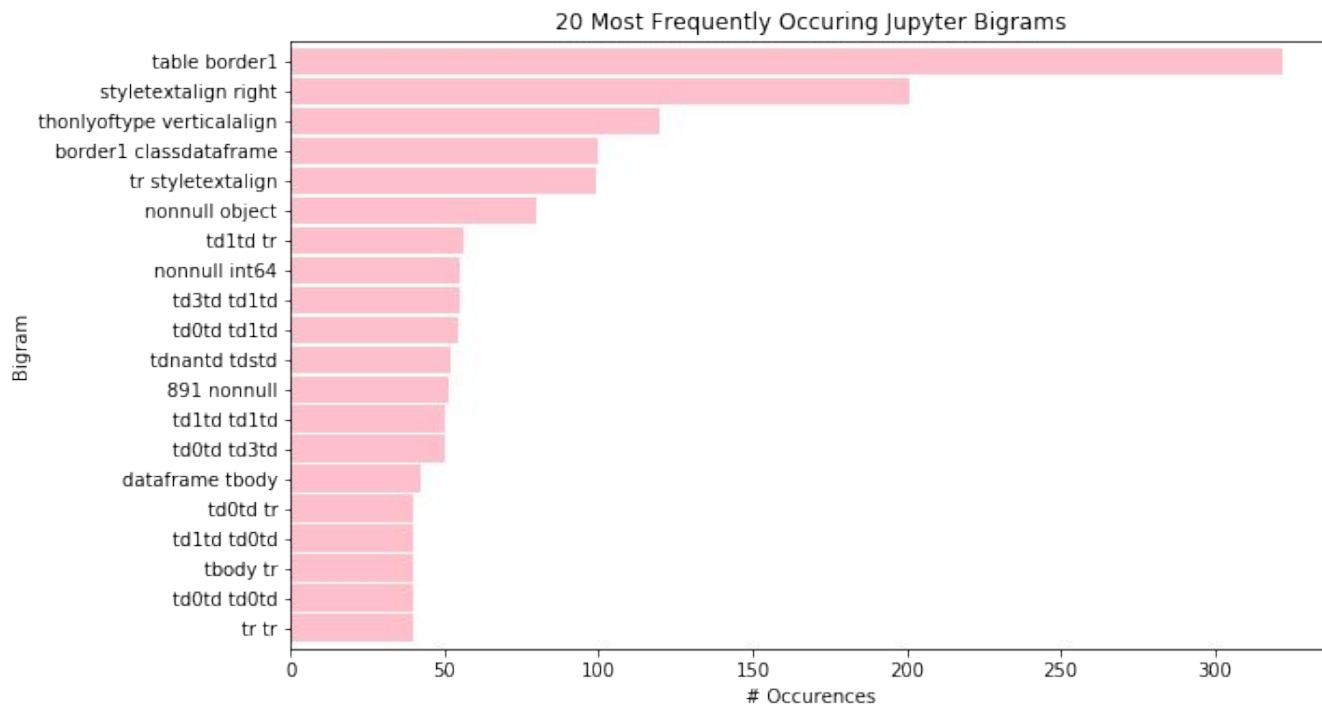
What does the data look like?

What are the languages of the most popular repositories for Titanic dataset?



Acquiring data!!!

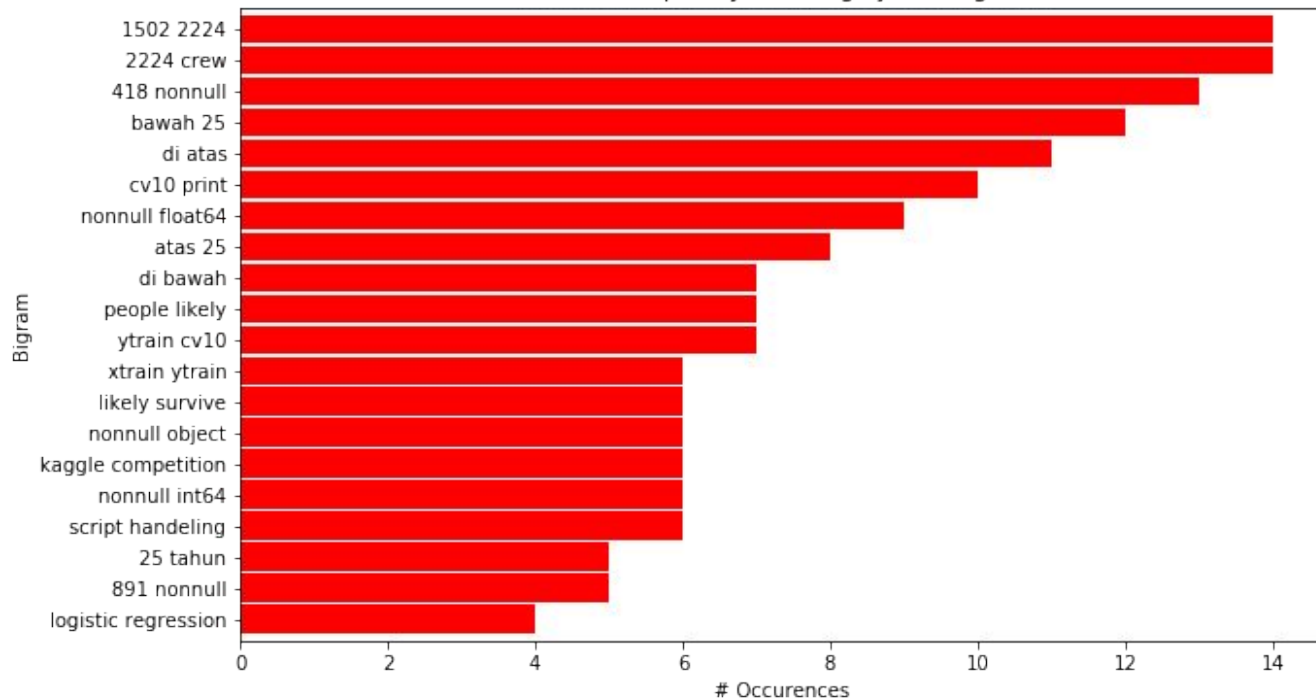
What does the data look like?



Exploring!!

What does the data look like?

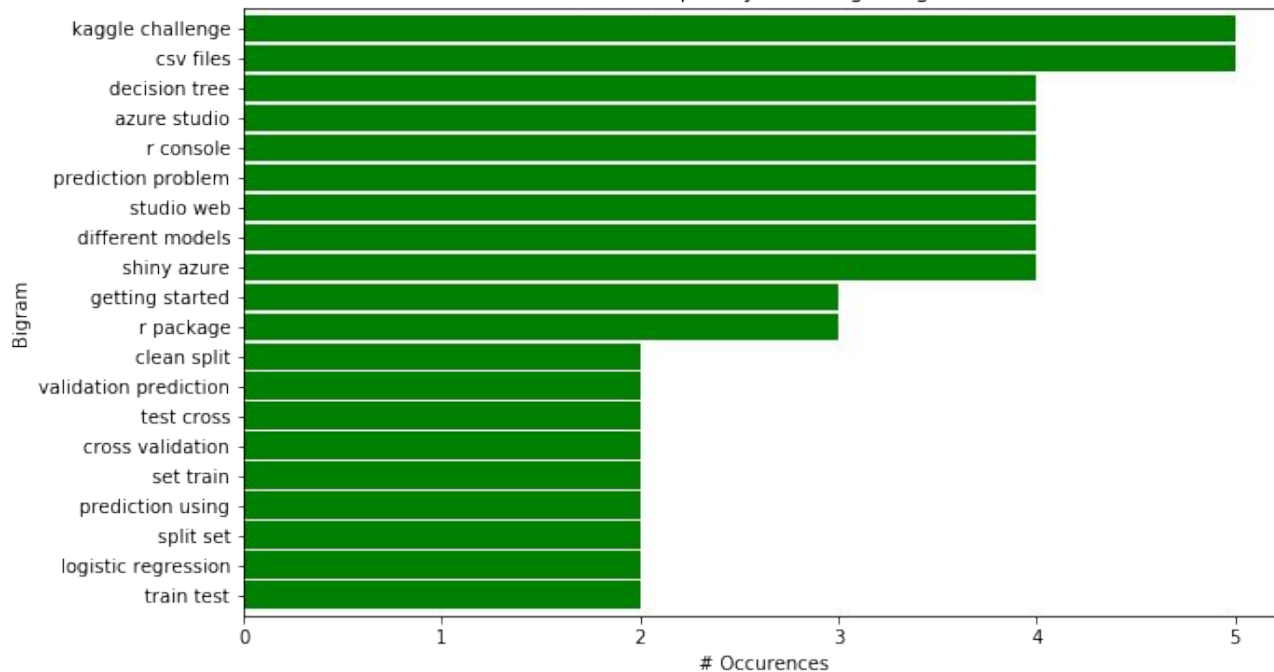
20 Most Frequently Occuring Python Bigrams



Exploring!!

What does the data look like?

20 Most Frequently Occurring R Bigrams



Exploring!!

Modeling

Hopes for our model

**Can we do better than a baseline?
(Accuracy)**

Baseline

.48

Logistic

Train: .53, Test: .49

Decision Tree

Train: .56, Test: .51



Our model

