STAR TREK CHARACTERS

Jennifer Gundersdorff

jenjohn27612@gmail.com

August 2024

Introduction

As a respite from my studies during the last several months, I determined to watch the various Star Trek series from start to finish. I have been amazed at the continuity of the characters throughout. I got curious about how we see these characters across the series continuum. Realizing it was time for me to do my first solo data analysis project, these characters seemed a natural choice. These are the questions I wanted to find out more about.

How many species are represented and what is the distribution?

How do the characters change across the series?

How are the various ranks (or non-ranks) of the characters represented?

How do the characters crossover to different Star Trek series?

The Dataset

Wikipedia had me covered as far as the data was concerned. A bit of web scraping with pandas and I had the basis for my analysis.

I scraped the Main and Recurring Character table as well as the Key Series table. Most guest appearances are not included, unless they were already a primary character.

I focused on the earlier series and decided to skip the films… for now.

The Original Series

The Next Generation

Deep Space 9

Voyager

Enterprise

more about the dataset

Series Abbreviation Key

Series	Title
TC	"The Cage" – original pilot episode 1966
TOS	The Original Series 1966-1969
TNG	The Next Generation 1987-1994
DS9	Deep Space 9 1993-1999
VOY	Voyager 1995-2001
ENT	Enterprise 2001-2005

Feature Information

Column	Description
Character	Character represented in Star Trek
Actor	Actor who played the given character(s)
Seasons	Seasons these characters appeared in the given series even if only briefly
Rank	Normalized rank as indicated in setting
Posting	Actual posting or home planet
Position	Normalized role for the characters
Species	Descriptive of entity or home planet
Season Count	Normalized count of season to reflect no more that one count per series season.

The Process

For the initial scraping and data cleaning I used a Jupyter notebook. ChatGPT4.0 assisted with Python code, troubleshooting and debugging. This included parsing the original 'Appearances' column into new 'Series' and 'Seasons' columns with new rows for each combination.

I used Excel to fine tune the more nuanced character designations that could not be readily parsed with Python. I normalized some of the similar rank and post information and corrected a season count issue in Excel as well.

I used Cognos Analytics to generate most of the visualizations. This went well though I do wish I could have modified the label size on the network visualization. I returned to the Jupyter notebook to create a series timeline with matplotlib copying the color codes used in Cognos for consistency.

cleaning challenge

Challenge

- Wikipedia had combined designations for multiple series and seasons in a single row in the 'Appearances', 'Ranking', 'Posting', and 'Position' columns.

Solution

- ChatGPT and I were able to find workable code to parse this in the 'Appearances' column. You can see specifics including debugging code in the Jupyter Notebook.
- For the 'Rank', 'Posting', and 'Position' columns, I switched to Excel and used text to columns then manually added new rows matching appropriate information. In some instances, character research was necessary to fully understand the correct assignments.

visualization adjustments

Challenge

- Upon starting the visualizations, I realized I had preserved too much nuanced information in some categories and that the season counts were sometimes off.
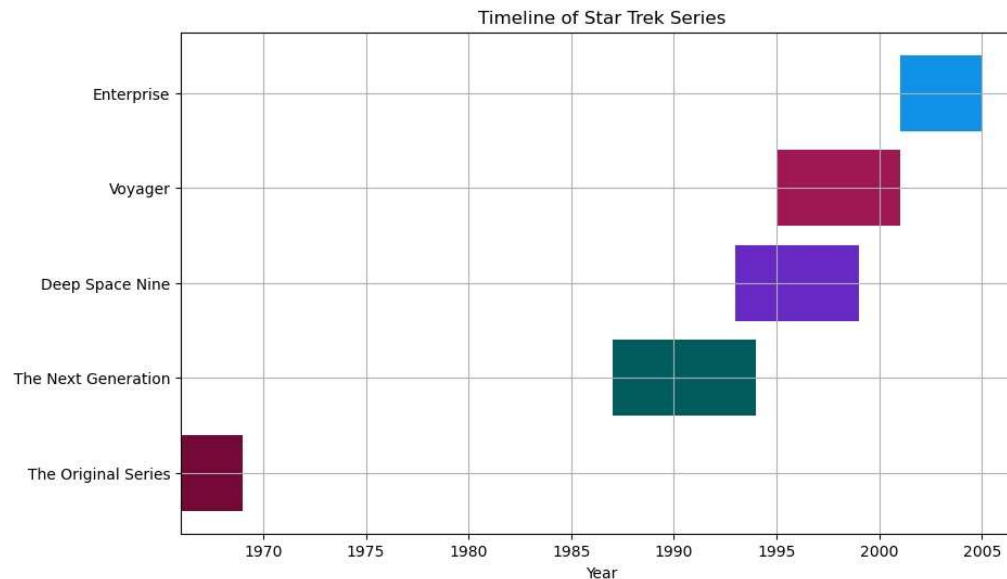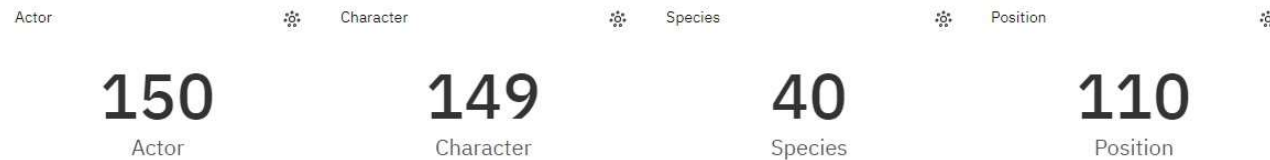
Solution

- I also normalized some of the rank, posting and position categories by eliminated descriptive words like 'provisional', 'acting' and 'resident'.
- The seasons had been overcounted due to multiple rows reflecting different ranks/postings within a single season. I fixed this up quickly in excel by crediting only partial seasons in the count column where that was the case.
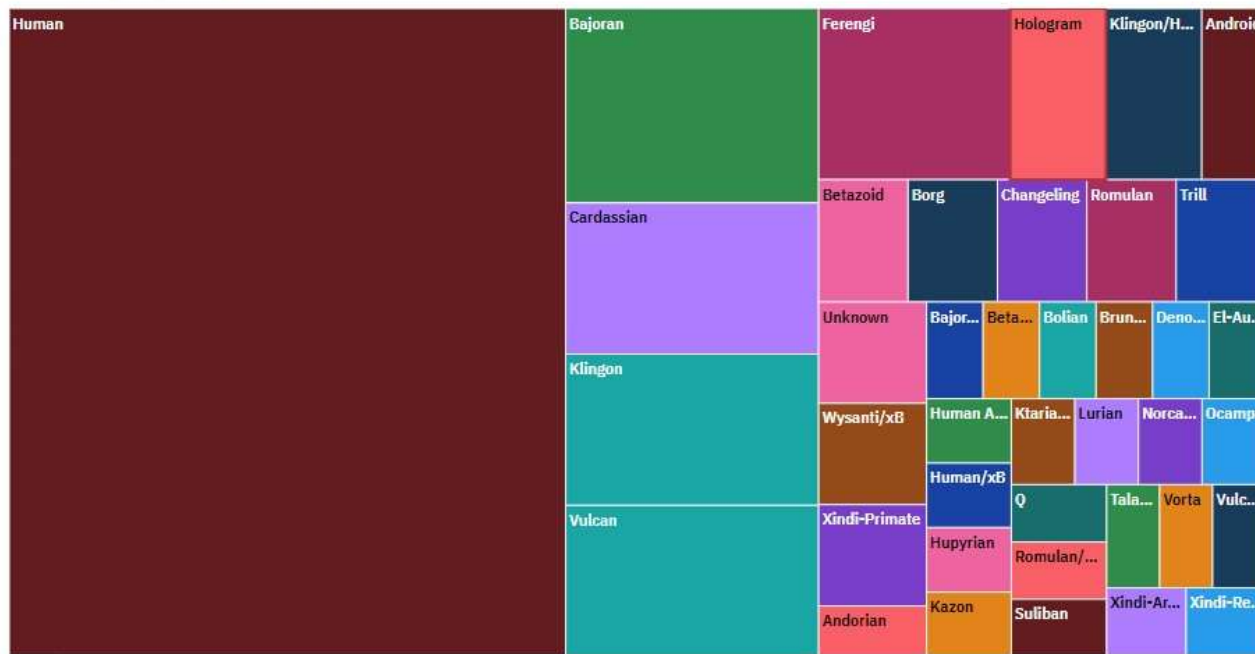
ANALYSIS

Star Trek Overview

Actor   Character   Species   Position

150   **149**   **40**   **110**
Actor   Character   Species   Position

Timeline of Star Trek Series

To start us off, here are some general statistics about the total number of distinct actors, characters, species and positions across our series of interest. This gives us the broad scope of just how many main and recurring characters were created, an idea of the consistency of the actors who portrayed them, and the range of diversity in species and occupation.

The timeline reminds us how this enduring franchise spanned the decades.

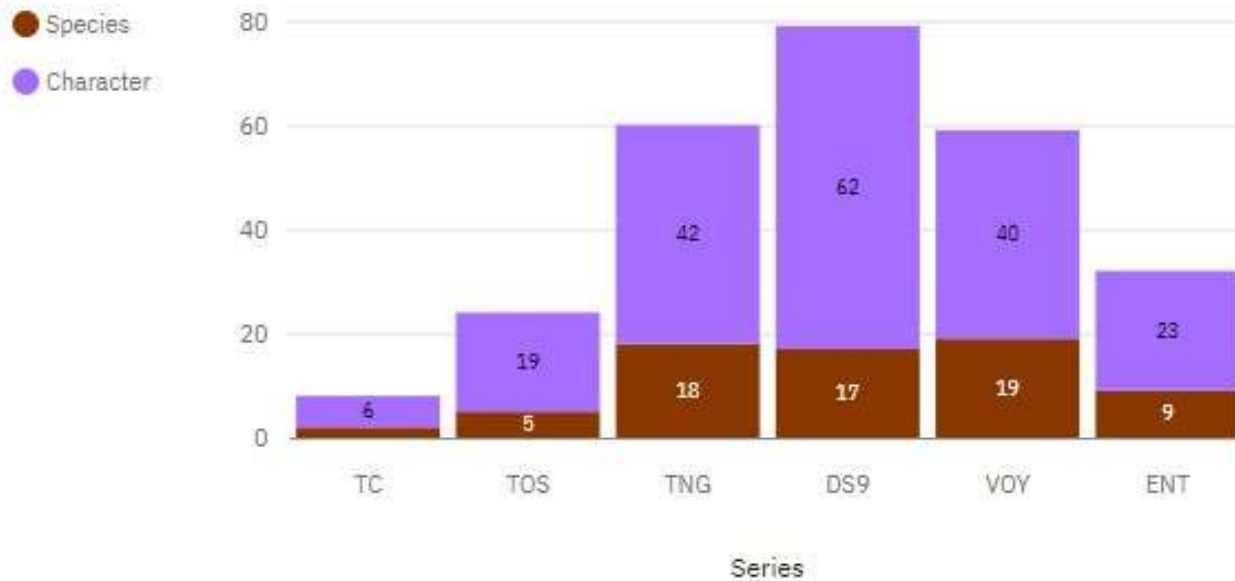# Star Trek Species

Character by Species hierarchy

As I watched the different episodes. (I am a bit into Deep Space 9 as I write this.) I could not help but notice more and more species, especially as the static wormhole in Deep Space 9 opens our explorers up to new worlds in the Gamma Quadrant.

Here is an overall look at the species of the main and recurring characters in our selected series. This treemap shows relative representation of the character count by species.

Star Trek Series Diversity

Species and Character by Series
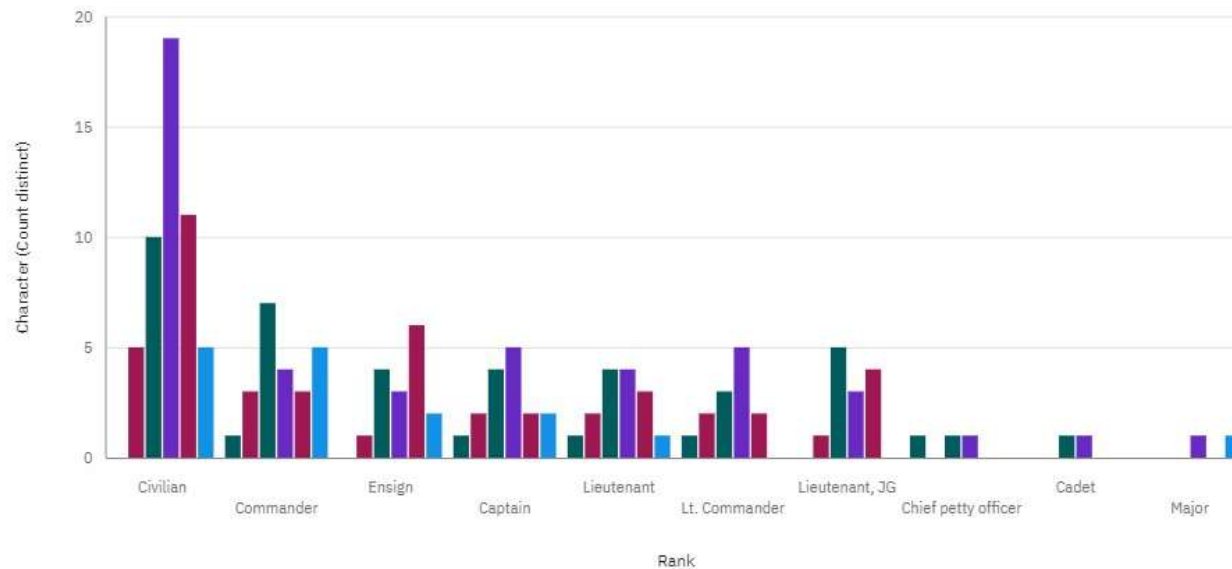

- Species
- Character

Here we can clearly see how the character and species diversity changed through the series. You can especially see how Deep Space 9 being at a crossroads in the Star Trek universe, had by far the greatest number of characters. Likely also the greatest number of species if guest characters had been included in the analysis.

The Next Generation and Voyager are pretty much tied for second place in terms of diversity thanks to explorations across seven seasons each.

Star Trek Character Ranks

Character by Rank colored by Series

Series
● TC ● TOS ● TNG ● DS9 ● VOY ● ENT

There are many different ranks indicated in our dataset. I focused on some of the key ranks so we can see how they are represented across the series.
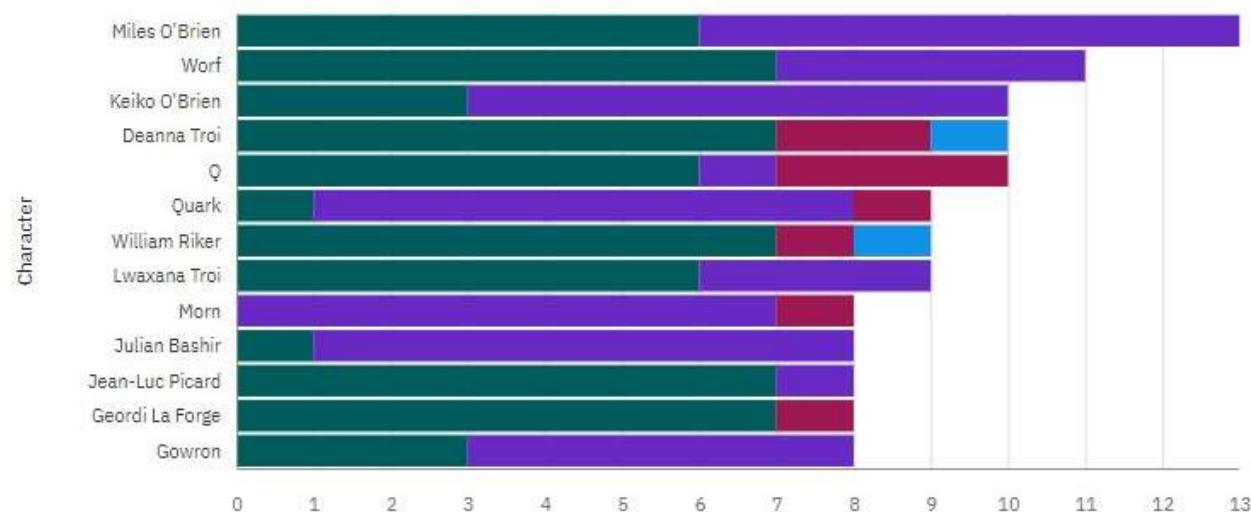
I am encouraged that so many civilians are included in the exploration of space, at last as imagined by science fiction writers.

Interestingly, the Chief Petty Officer indicated for TNG and DS9 is none other than Miles O'Brien the character with the second most appearances across the franchise.[1]

Star Trek Top Characters by Series

Season Count by Character colored by Series

Series
● TNG ● DS9 ● VOY ● ENT

Here we have Miles at the top of the list for number of seasons followed by Worf, who is the Star Trek character with the most appearances.[2] To keep this project to a reasonable amount of time I focused on seasons rather than episodes, but they would certainly be fertile ground for further exploration!

Deanna Troi, Q, Quark, and Riker are notable for appearing in three different series, though some of those are only brief appearances.

Conclusions

The Star Trek writers have done a phenomenal job of keeping key characters relevant and faithfully represented across its many series.

Species diversity is a cornerstone of the Star Trek universe, consistently and authentically portrayed throughout the franchise.

The range of characters from ensigns to admirals, civilians to religious leaders are woven into the storyline showing interactions inclusive of all areas of life.

I would love to expand this analysis to include greater detail, possibly to the episode level, and to include the films and additional series with more interactive visualizations.

Resources

Data Sources

[Wikipedia – List of Star Trek Characters](#)

[1] [Miles O'Brien anecdote](#)

[2] [Worf anecdote](#)

Tools

Jupyter notebook

ChatGPT 4.0

Cognos Analytics

Python Libraries

- pandas
- re
- matplotlib

Project Links

[GitHub repository](#)

[Jupyter notebook](#)

[Final Excel data](#)

APPENDIX

Character Network Map

I am unsatisfied with the label size of character names as generated in Cognos. Yet, I decided to include if for those interested in zooming in on screen to see how our characters are involved in the different series.

Star Trek Character Network Map