

# NASA Needs Fresh Meat

## *An Outsider's Analysis of NASA's Current Astronaut Roster*

by Graham Ganssle, Ph.D.

As a scientist (and an adrenaline junkie), getting blasted off the planet by 150,000 kilos of liquid oxygen and kerosene sounds awesome to me, so I submitted an application to NASA to join the new cohort of astronaut candidates. However, judging by historical data, it's likely applicants like myself have a chance of acceptance on the order of forty thousandths of a percent. Those odds are pretty thin, but the home run of acceptance isn't the only derived benefit of the application process. It's been a great pleasure to meet fellow astronaut candidate applicants and former/current employees of NASA. These people are the brightest, hardest working people in the US, so how does NASA choose between them and what skills is the current NASA astronaut roster missing?

During my whimsical Googlings about NASA's current astronaut corps, I found a page with their biographical information. I hypothesized it'd be possible to answer the above questions using this information, so I banged on my keyboard for a while until I learned Python, how to webscrape PDFs, and the intricacies of the human-language-data / machine-language-data interface. I'm not sure I came up with any interesting answers, but the simple analysis I did, and the cleaned data itself, isn't available anywhere else. So for the benefit of NASA (who probably already has this information somewhere!) and the eight or so applicants who have the Right Stuff, I hope this information helps you out in some small way. For the rest of us, I hope this project is the starting point for a conversation between likeminded data nerds, and can be a place to meet NASA interested scientists/engineers/soldiers/teachers/pilots/etc of the highest caliber.

The project can be accessed here: <https://github.com/gganssle/astroBio> Please comment, suggest, add to, or remove from this as you see fit. But above all, introduce yourself!

### **Methods**

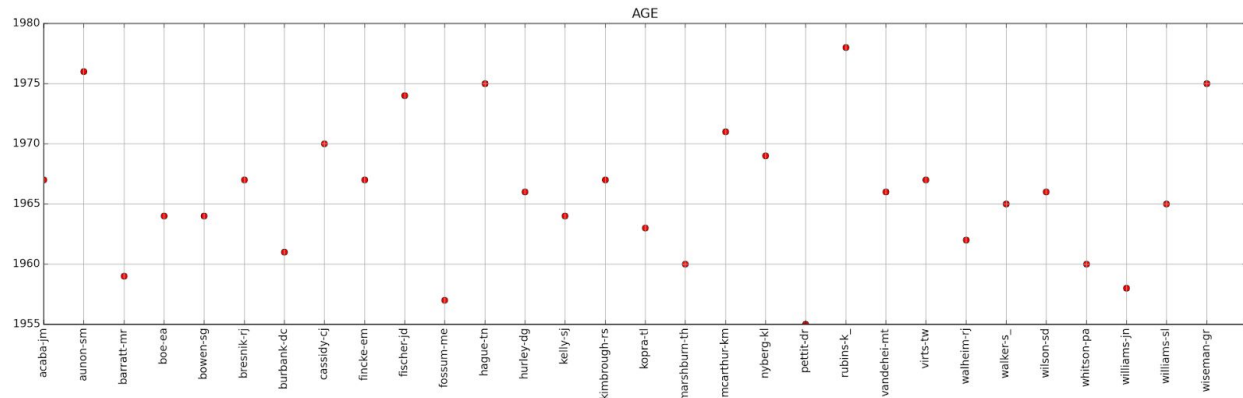
I used Mathematica to webscrape the raw data from the interwebs, and then cleaned the PDF data to txt with pdftotext.com. I wrote a Python program to do a rough clean on the textual data, and then a Python script in a Jupyter notebook to do the analysis and make things look pretty. Matplotlib was instrumental in presenting the data in a meaningful way. Initially I planned to use the excellent NLTK (Natural Language Toolkit), but the analysis within wasn't sophisticated enough to warrant its use.

### **Results**

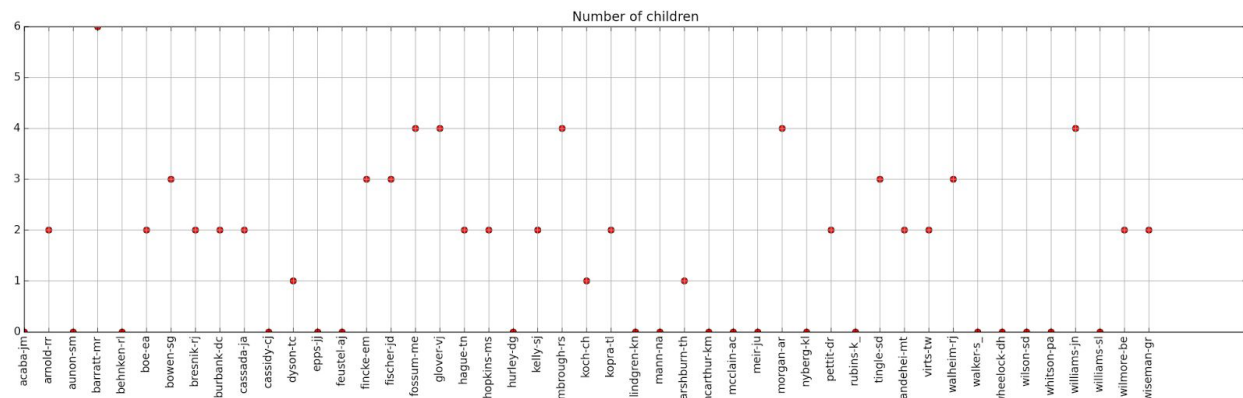
All of these figures can be found full size for detail in the open source repository:

<https://github.com/gganssle/astroBio>

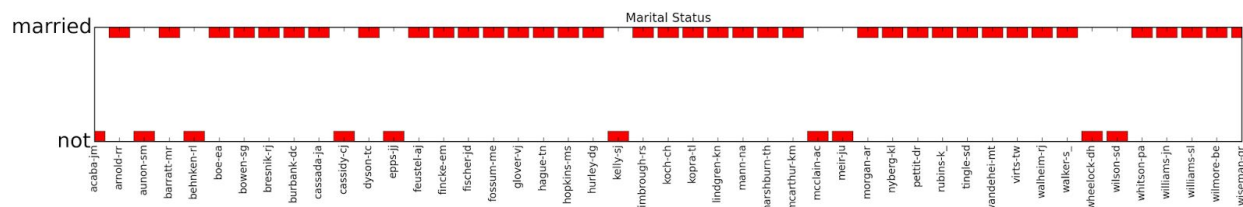
Only 83% of the biographies had age data, but of the data which was populated, the average age of the current astronaut corps is 50 years old. The figure below shows all of the age data populated in the biographies.



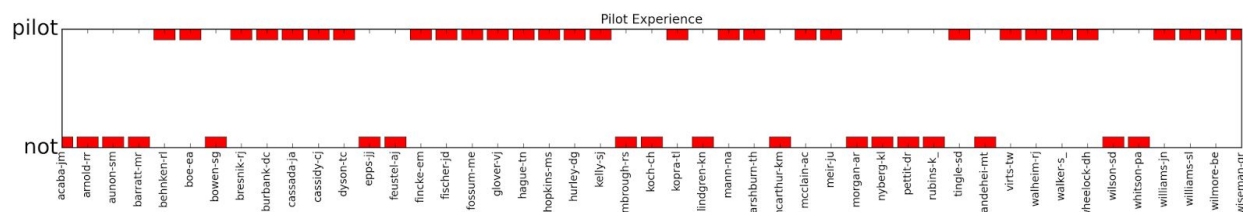
As expected, the average number of children the astronauts have is near the national average.



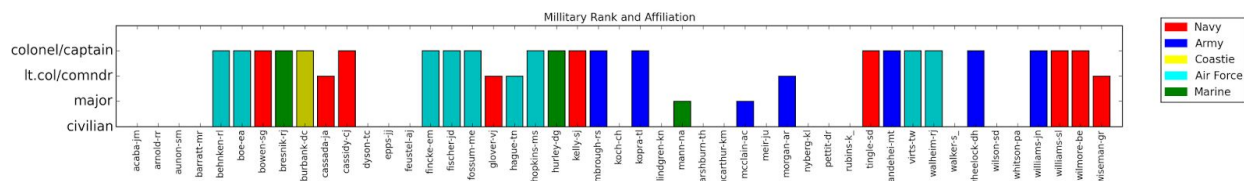
No information was disclosed in the biographies about explicitly unmarried astronauts, so this analysis assumes if not stated explicitly “married” the astronauts were counted unmarried.



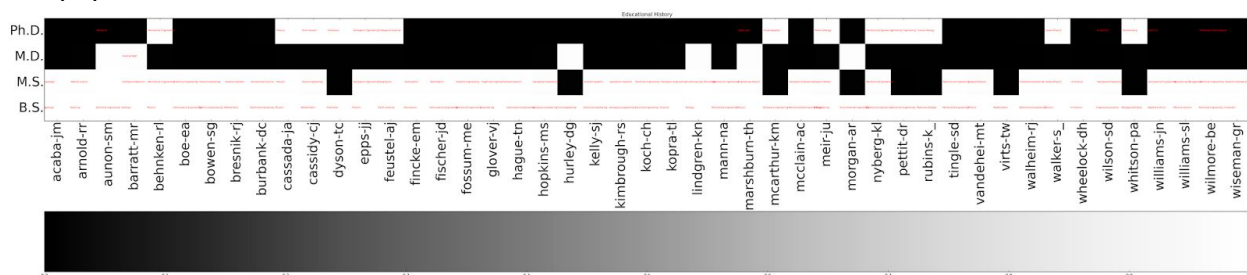
The next figure discloses flight experience, nondiscriminating of military or private experience.



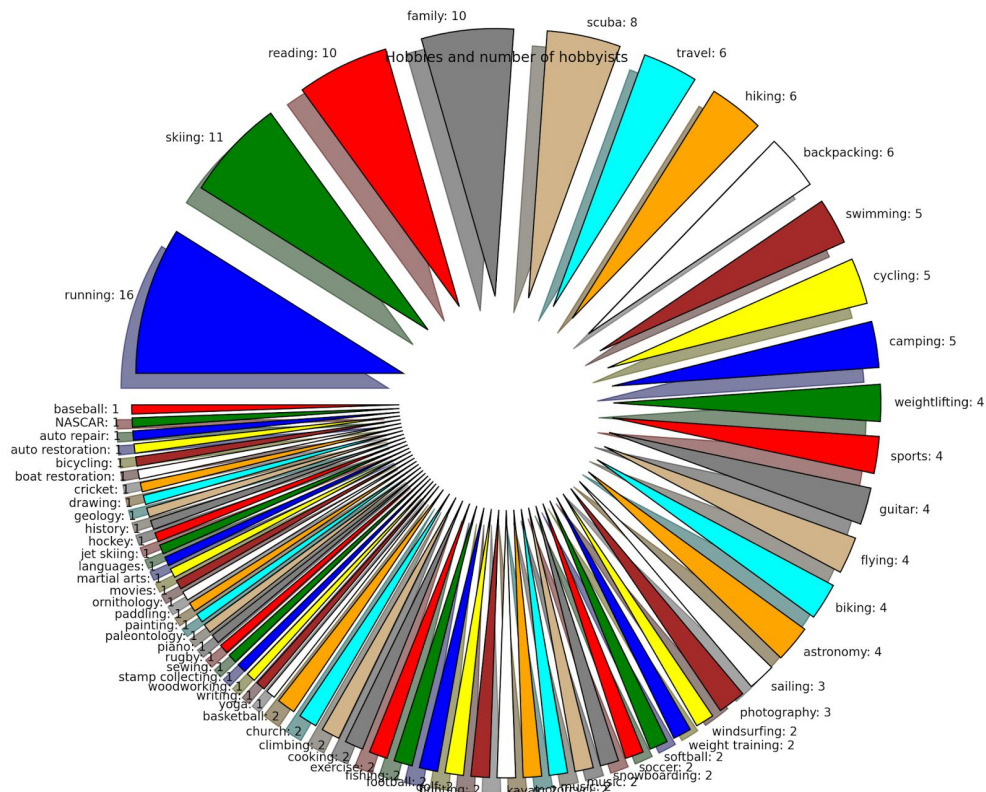
The following figure describes each astronaut's military affiliation and rank, with service branch in color and rank as bar height.



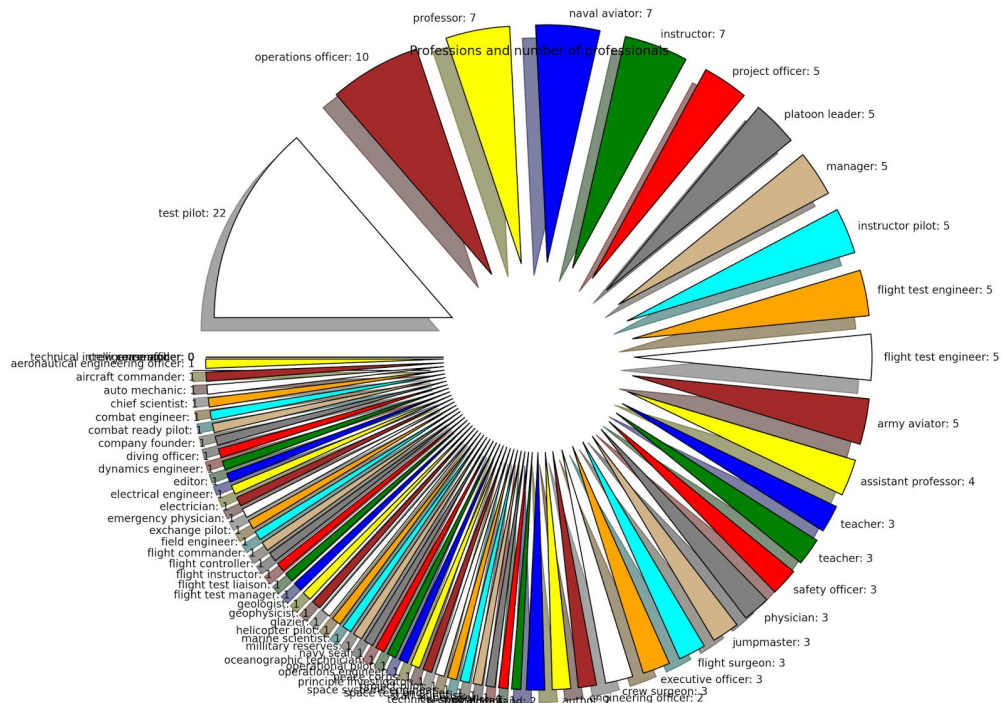
The figure below (again, see <https://github.com/gganssle/astroBio> for a full size version where you can read the details) shows the astronauts' educational history. A white square means they have the degree, a black square means they do not. The red labels indicate the specialty of each degree. The process was limited by syntactical constraints, as you'll notice by a few mis-populations.



Hooray! Pie charts! The next figure is a representation of how many astronauts share the same personal interests and hobbies. The numbers next to the tags are how many astronauts share the hobby.



Work history is represented in the next figure. The number next to the tag is the number of astronauts who've had the same profession.



## Conclusions