R &| Data Science

Dr. Jennifer Piscionere

Somewhere in Swinburne

How I Got An (Academic) Data Science Job

I knew a guy

Coding in R

Because All the Job Adverts Mention It

100%

Percent of Us Who Will Be Looking for a Job

Astronomer

Quant

Data Scientist

CEO of Household

General of Tiny People

Emperor of ‘I Just Need Some Time’

Marketing Executive

‘Stop Bad Hombres from Getting into This Data Set’ People

Start Up Maven

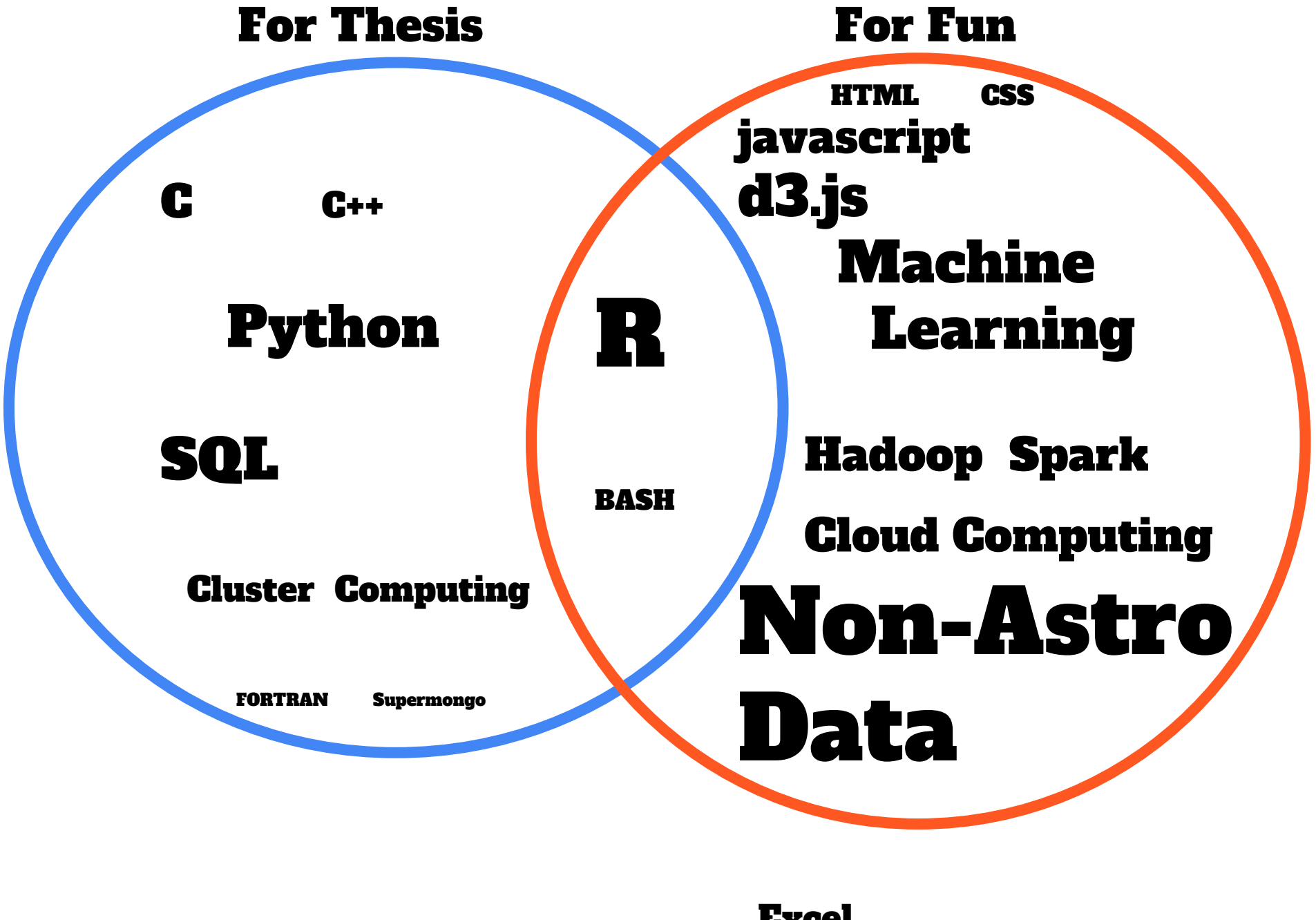
‘I Actually Have A Creative Side Too’ Website Designer

‘I Want to Cure Cancer’ BioInformatrix

Siren Telling the Government How To Spend Money on Science (me)

100%

Percent of Us Who Will Get a Job


For Thesis
For Fun
C
C++
Python
R
SQL
BASH
Cluster Computing
FORTRAN
Supermongo
HTML
CSS
javascript
d3.js
Machine Learning
Hadoop
Spark
Cloud Computing
Non-Astro Data
Excel

Build My Own Website

HTML CSS
javascript
d3.js

Workshop

Hadoop
Spark

Kaggle Tutorial

Machine
Learning

Data Science Fellowship

Non-Astro Data

Cloud Computing

Aside On Non-Astro Data

Second Question You Will Be Asked in a Job Interview:

Why Do You Want *this* Job?

Not ***A*** Job, Not ***Any*** Job,
Not ***Dear God Please I'll Take Any Job at this Point*** Job

Top Tip: Apply for Jobs You Want

‘Data Science’ Is Nebulous.
It helps to know what you want you want to do with your skills.

Working with Data Outside of Astronomy Can Help.

And look rreeeeeeeeaaaaaaaaallllllllllyyyyyyyyyyyyyyyyyy good on a resume and in the interview.

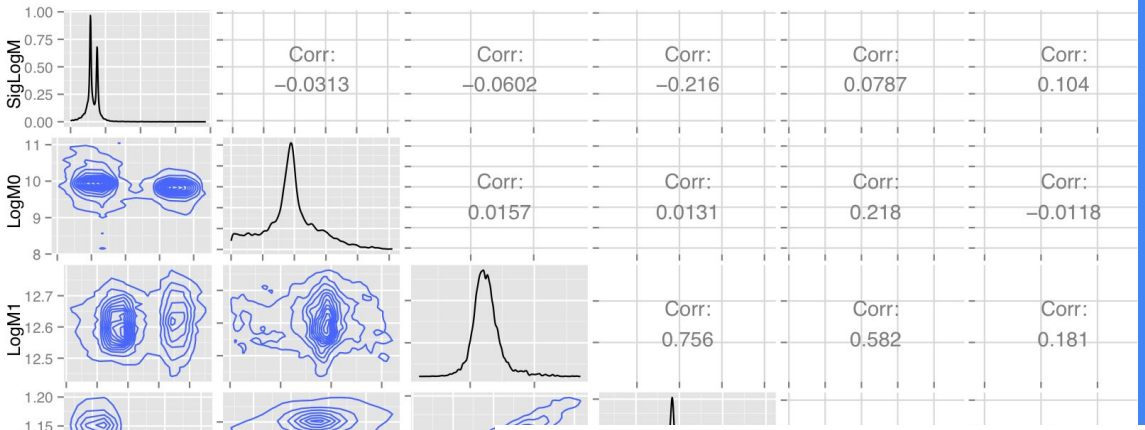
Kaggle is One Solution

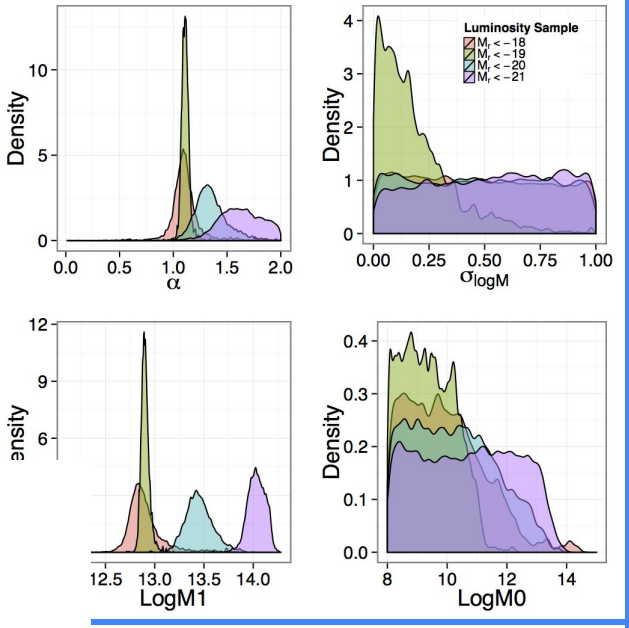
"The best place to discover and seamlessly analyze open data"

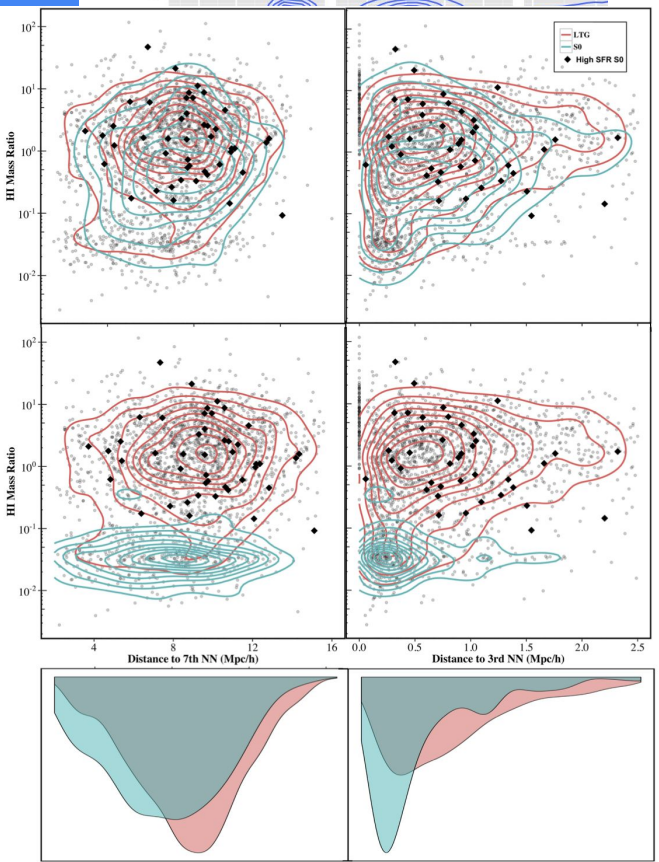
1. Go to https://www.kaggle.com/
2. Sign Up
3. Play
4. Try https://www.kaggle.com/c/titanic#tutorials
5. Add Data Camp Completion Certificate to LinkedIn Profile*

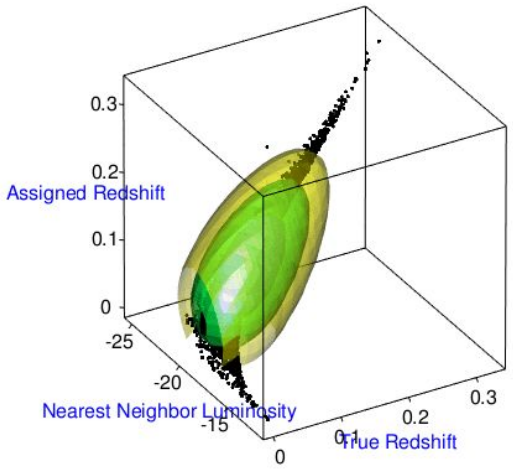
**You Have A LinkedIn Profile right?*

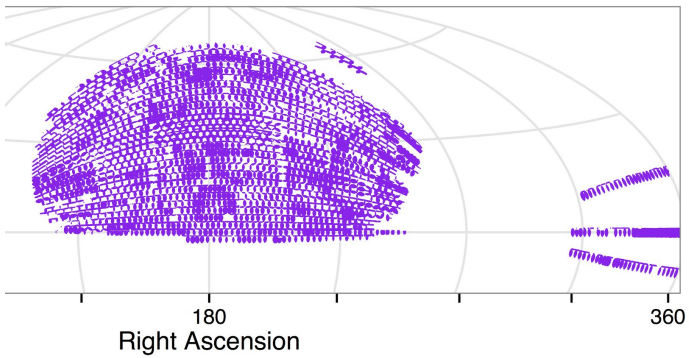
Time To R

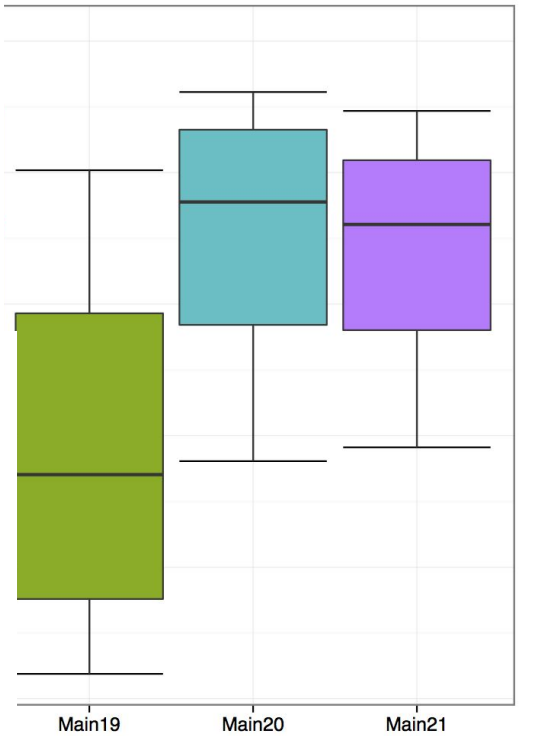

SigLogM
LogM0
LogM1
Alpha
Corr: −0.0313
Corr: −0.0602
Corr: −0.216
Corr: 0.0787
Corr: 0.104
Corr: 0.0157
Corr: 0.0131
Corr: 0.218
Corr: −0.0118
Corr: 0.756
Corr: 0.582
Corr: 0.181


Density
α
Luminosity Sample
M_r < −18
M_r < −19
M_r < −20
M_r < −21
σ_logM
LogM1
LogM0


LTG
S0
High SFR S0
HI Mass Ratio
Distance to 7th NN (Mpc/h)
Distance to 3rd NN (Mpc/h)


Assigned Redshift
Nearest Neighbor Luminosity
True Redshift


180
360
Right Ascension


Main19
Main20
Main21

1. https://cran.r-project.org/mirrors.html

*you can use macports too, but I just spend 1.5hrs updating all my R paths because macports doesn't install R in the right place which makes it hard to iPython Notebooks so consider yourself warned

A Few Ways To R

1. Run the GUI
2. Run on the Command Line
    a. Type ‘R’
3. Run in iPython Notebooks
    a. https://github.com/IRkernel/IRkernel

R Script:

http://jpiscionere.github.io/CodeReview.r

Data:

http://jpiscionere.github.io/HOPCAT2005TabSep.csv