What is EDA Notes

Lesson 1:

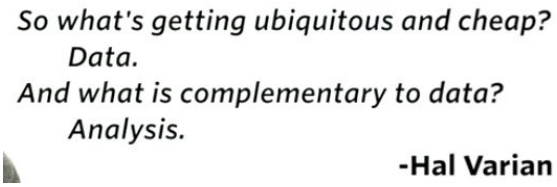
- Can use google trends to see what's popular right now.
    - https://www.google.com/trends/

Lesson 2:

- Look at the correlation!
    - https://www.google.com/trends/explore?q=vegan,healthy,diet,gym,health

---

Lesson 3:



- **Exploratory Data Analysis (EDA)** is the examination of data and relationships among variables, through both numerical and graphical methods.
    - Often takes place before more formal, more rigorous statistical analysis.
    - Often first part of larger process.
    - Important line of defence against bad data.

Lesson 5:

- **Exploratory Data Analysis (EDA)** is an approach toward understanding data using visualizations and statistical tools.
- **Goals:**
    - Understand the distribution of the variables
        - E.g. histograms or scatter plots
    - Assess and validate assumptions on which future inferences will be based.
        - E.g. which variables are normally distributed or whether a variable is biased towards a particular value.
    - We want to understand the data before forming intelligent hypothesis.
- **Ultimately we are developing intuition about our dataset and how it came into existence.**
    - By examining our data we can generate better hypothesis, determine which variables have the most predictive power and select appropriate statistical tools to build our predictive models.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4

Lesson 6:

- Good website for data visualisation
    - http://flowingdata.com/

Lesson 7:

- Always be curious and sceptical when looking at data.
- When conducting an EDA, we want to test our intuitions about the dataset and develop new intuitions.

On average people should be more skeptical when they see numbers. They should be more willing to play around with the data themselves.

-Nate Silver

Video: Our Approach For This Course

At 0:40, Dean discusses testing intuitions about a data set. One example to consider from the Facebook data set is that total likes by a Facebook user should be greater or equal to mobile likes by that user.