DALLAS FORMULA RACING

Week 1

Intro to Data Analysis

Agenda

- Course Overview
- Intro to Data Analysis
- Data Cleaning
- Live Walkthrough
- Updates / Reminders

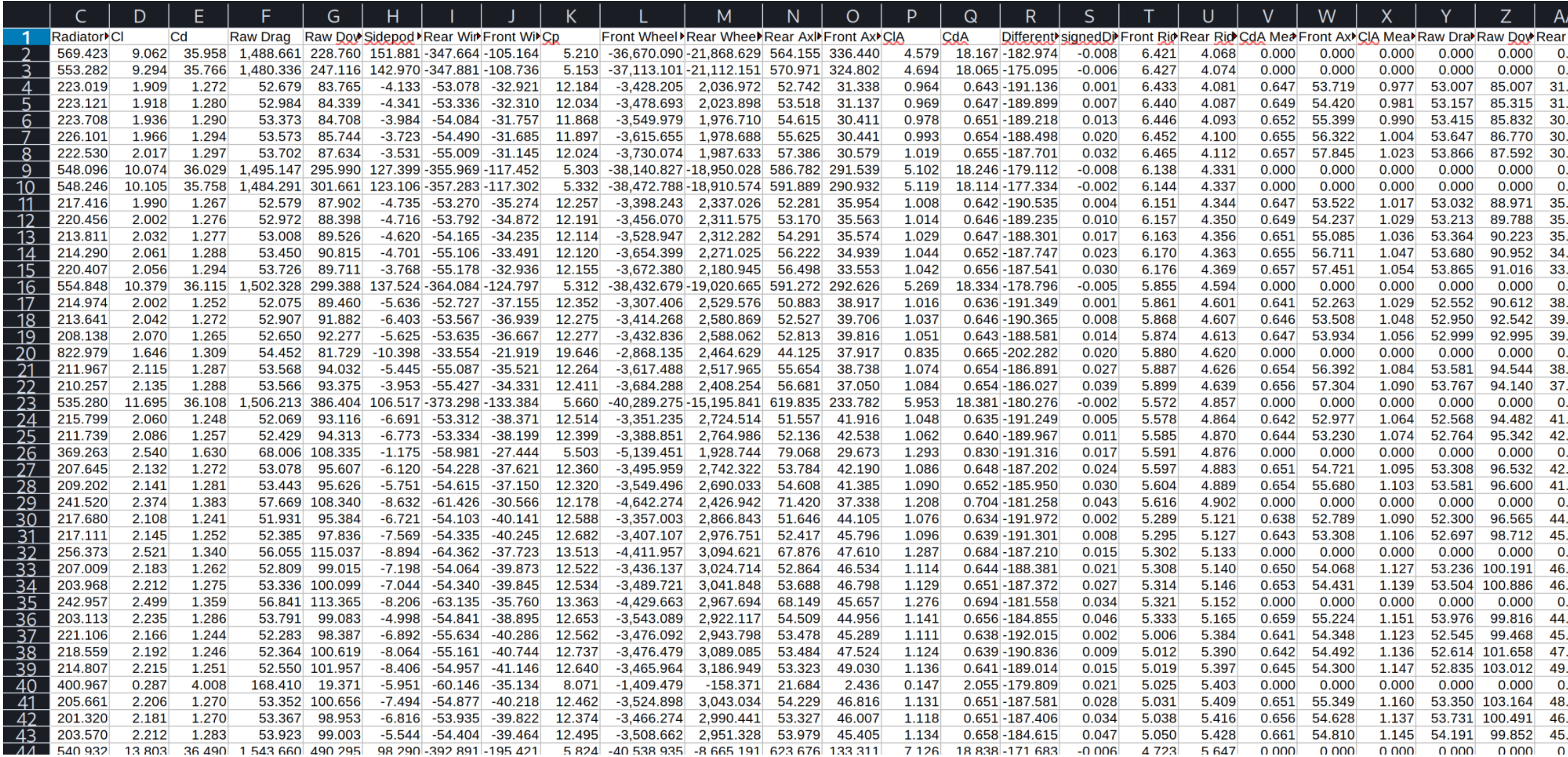
Course Plan (Dates subject to change for midterms)

- Intro to Data Analysis (Sep 23, 1PM ECSW 1.315):
  - Overview on how to think about data
  - What is Data Analysis? What is the process?
  - Live demonstration of the whole process
  - Applying this to writing testing plans and testing the car
- Applied Data Analysis (Sep 30, 1PM ECSW 1.315):
  - Applying this lecture to work through a whole analysis process
  - Workshop day with live example
  - Work through cleaning, visualization, analysis, correlation, and validation
  - Will use real data off of the car
  - Emphasis on correlation and validation with simulations
- Proces UTA Autocross Data (Oct 14th, 1PM ECSW 1.315)
  - Workshop day
  - Applying past lectures to real life
  - Apply what you learn in your sub team meetings to help find trends
  - Hopefully take these skills to use with your own sub team

DALLAS FORMULA RACING

Introduction

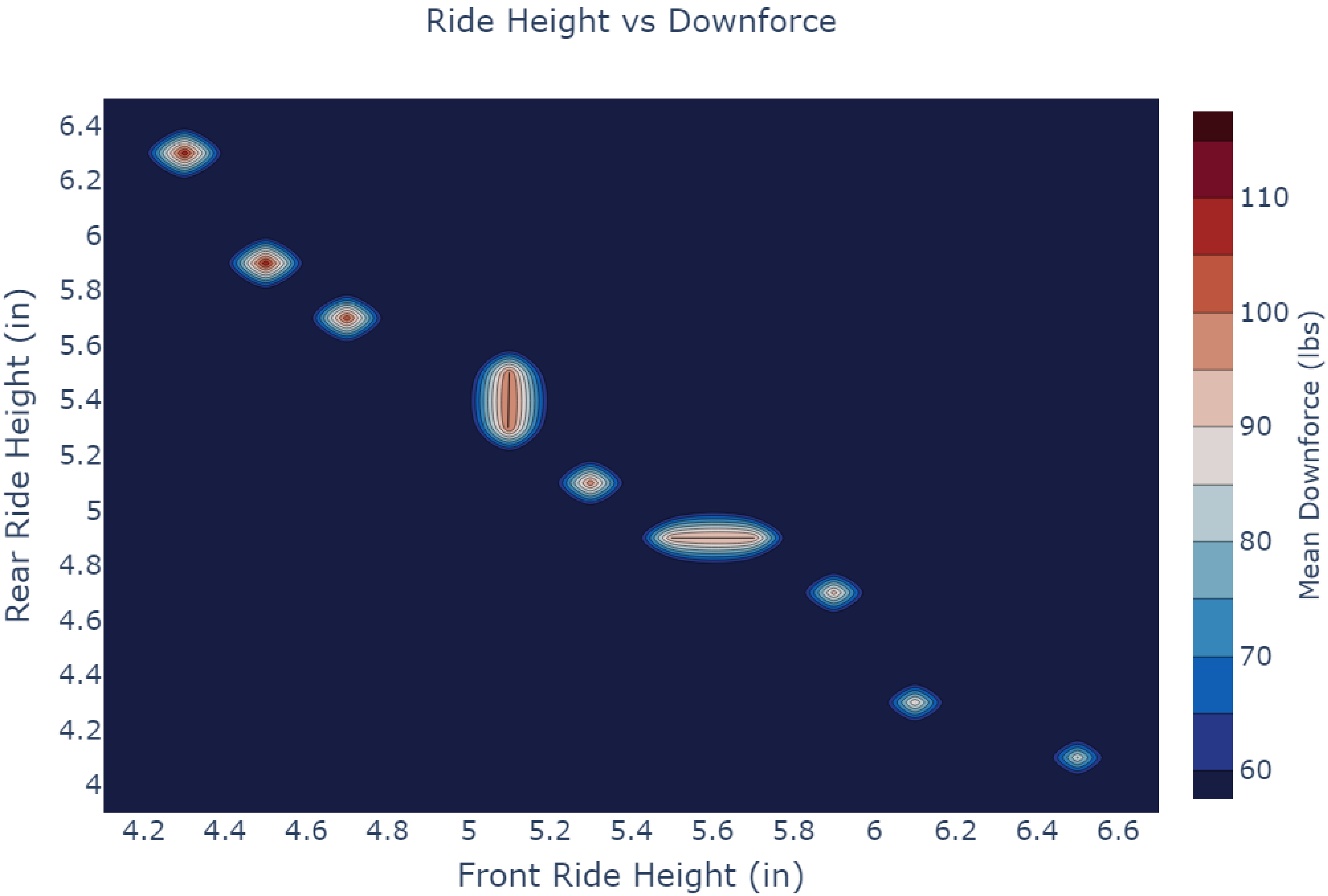
- How do we go from this:



	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
1	Radiator	Cl	Cd	Raw Drag	Raw Dow	Sidepod	Rear Wir	Front Wi	Cp	Front Wheel	Rear Whee	Rear Axl	Front Ax	ClA	CdA	Different	signedDi	Front Rid	Rear Rid	CdA Mea	Front Ax	ClA Mea	Raw Dra	Raw Dov	Rear
2	569.423	9.062	35.958	1,488.661	228.760	151.881	-347.664	-105.164	5.210	-36,670.090	-21,868.629	564.155	336.440	4.579	18.167	-182.974	-0.008	6.421	4.068	0.000	0.000	0.000	0.000	0.000	0.
3	553.282	9.294	35.766	1,480.336	247.116	142.970	-347.881	-108.736	5.153	-37,113.101	-21,112.151	570.971	324.802	4.694	18.065	-175.095	-0.006	6.427	4.074	0.000	0.000	0.000	0.000	0.000	0.
4	223.019	1.909	1.272	52.679	83.765	-4.133	-53.078	-32.921	12.184	-3,428.205	2,036.972	52.742	31.338	0.964	0.643	-191.136	0.001	6.433	4.081	0.647	53.719	0.977	53.007	85.007	31.
5	223.121	1.918	1.280	52.984	84.339	-4.341	-53.336	-32.310	12.034	-3,478.693	2,023.898	53.518	31.137	0.969	0.647	-189.899	0.007	6.440	4.087	0.649	54.420	0.981	53.157	85.315	31.
6	223.708	1.936	1.290	53.373	84.708	-3.984	-54.084	-31.757	11.868	-3,549.979	1,976.710	54.615	30.411	0.978	0.651	-189.218	0.013	6.446	4.093	0.652	55.399	0.990	53.415	85.832	30.
7	226.101	1.966	1.294	53.573	85.744	-3.723	-54.490	-31.685	11.897	-3,615.655	1,978.688	55.625	30.441	0.993	0.654	-188.498	0.020	6.452	4.100	0.655	56.322	1.004	53.647	86.770	30.
8	222.530	2.017	1.297	53.702	87.634	-3.531	-55.009	-31.145	12.024	-3,730.074	1,987.633	57.386	30.579	1.019	0.655	-187.701	0.032	6.465	4.112	0.657	57.845	1.023	53.866	87.592	30.
9	548.096	10.074	36.029	1,495.147	295.990	127.399	-355.969	-117.452	5.303	-38,140.827	-18,950.028	586.782	291.539	5.102	18.246	-179.112	-0.008	6.138	4.331	0.000	0.000	0.000	0.000	0.000	0.
10	548.246	10.105	35.758	1,484.291	301.661	123.106	-357.283	-117.302	5.332	-38,472.788	-18,910.574	591.889	290.932	5.119	18.114	-177.334	-0.002	6.144	4.337	0.000	0.000	0.000	0.000	0.000	0.
11	217.416	1.990	1.267	52.579	87.902	-4.735	-53.270	-35.274	12.257	-3,398.243	2,337.026	52.281	35.954	1.008	0.642	-190.535	0.004	6.151	4.344	0.647	53.522	1.017	53.032	88.971	35.
12	220.456	2.002	1.276	52.972	88.398	-4.716	-53.792	-34.872	12.191	-3,456.070	2,311.575	53.170	35.563	1.014	0.646	-189.235	0.010	6.157	4.350	0.649	54.237	1.029	53.213	89.788	35.
13	213.811	2.032	1.277	53.008	89.526	-4.620	-54.165	-34.235	12.114	-3,528.947	2,312.282	54.291	35.574	1.029	0.647	-188.301	0.017	6.163	4.356	0.651	55.085	1.036	53.364	90.223	35.
14	214.290	2.061	1.288	53.450	90.815	-4.701	-55.106	-33.491	12.120	-3,654.399	2,271.025	56.222	34.939	1.044	0.652	-187.747	0.023	6.170	4.363	0.655	56.711	1.047	53.680	90.952	34.
15	220.407	2.056	1.294	53.726	89.711	-3.768	-55.178	-32.936	12.155	-3,672.380	2,180.945	56.498	33.553	1.042	0.656	-187.541	0.030	6.176	4.369	0.657	57.451	1.054	53.865	91.016	33.
16	554.848	10.379	36.115	1,502.328	299.388	137.524	-364.084	-124.797	5.312	-38,432.679	-19,020.665	591.272	292.626	5.269	18.334	-178.796	-0.005	5.855	4.594	0.000	0.000	0.000	0.000	0.000	0.
17	214.974	2.002	1.252	52.075	89.460	-5.636	-52.727	-37.155	12.352	-3,307.406	2,529.576	50.883	38.917	1.016	0.636	-191.349	0.001	5.861	4.601	0.641	52.263	1.029	52.552	90.612	38.
18	213.641	2.042	1.272	52.907	91.882	-6.403	-53.567	-36.939	12.275	-3,414.268	2,580.869	52.527	39.706	1.037	0.646	-190.365	0.008	5.868	4.607	0.646	53.508	1.048	52.950	92.542	39.
19	208.138	2.070	1.265	52.650	92.277	-5.625	-53.635	-36.667	12.277	-3,432.836	2,588.062	52.813	39.816	1.051	0.643	-188.581	0.014	5.874	4.613	0.647	53.934	1.056	52.999	92.995	39.
20	822.979	1.646	1.309	54.452	81.729	-10.398	-33.554	-21.919	19.646	-2,868.135	2,464.629	44.125	37.917	0.835	0.665	-202.282	0.020	5.880	4.620	0.000	0.000	0.000	0.000	0.000	0.
21	211.967	2.115	1.287	53.568	94.032	-5.445	-55.087	-35.521	12.264	-3,617.488	2,517.965	55.654	38.738	1.074	0.654	-186.891	0.027	5.887	4.626	0.654	56.392	1.084	53.581	94.544	38.
22	210.257	2.135	1.288	53.566	93.375	-3.953	-55.427	-34.331	12.411	-3,684.288	2,408.254	56.681	37.050	1.084	0.654	-186.027	0.039	5.899	4.639	0.656	57.304	1.090	53.767	94.140	37.
23	535.280	11.695	36.108	1,506.213	386.404	106.517	-373.298	-133.384	5.660	-40,289.275	-15,195.841	619.835	233.782	5.953	18.381	-180.276	-0.002	5.572	4.857	0.000	0.000	0.000	0.000	0.000	0.
24	215.799	2.060	1.248	52.069	93.116	-6.691	-53.312	-38.371	12.514	-3,351.235	2,724.514	51.557	41.916	1.048	0.635	-191.249	0.005	5.578	4.864	0.642	52.977	1.064	52.568	94.482	41.
25	211.739	2.086	1.257	52.429	94.313	-6.773	-53.334	-38.199	12.399	-3,388.851	2,764.986	52.136	42.538	1.062	0.640	-189.967	0.011	5.585	4.870	0.644	53.230	1.074	52.764	95.342	42.
26	369.263	2.540	1.630	68.006	108.335	-1.175	-58.981	-27.444	5.503	-5,139.451	1,928.744	79.068	29.673	1.293	0.830	-191.316	0.017	5.591	4.876	0.000	0.000	0.000	0.000	0.000	0.
27	207.645	2.132	1.272	53.078	95.607	-6.120	-54.228	-37.621	12.360	-3,495.959	2,742.322	53.784	42.190	1.086	0.648	-187.202	0.024	5.597	4.883	0.651	54.721	1.095	53.308	96.532	42.
28	209.202	2.141	1.281	53.443	95.626	-5.751	-54.615	-37.150	12.320	-3,549.496	2,690.033	54.608	41.385	1.090	0.652	-185.950	0.030	5.604	4.889	0.654	55.680	1.103	53.581	96.600	41.
29	241.520	2.374	1.383	57.669	108.340	-8.632	-61.426	-30.566	12.178	-4,642.274	2,426.942	71.420	37.338	1.208	0.704	-181.258	0.043	5.616	4.902	0.000	0.000	0.000	0.000	0.000	0.
30	217.680	2.108	1.241	51.931	95.384	-6.721	-54.103	-40.141	12.588	-3,357.003	2,866.843	51.646	44.105	1.076	0.634	-191.972	0.002	5.289	5.121	0.638	52.789	1.090	52.300	96.565	44.
31	217.111	2.145	1.252	52.385	97.836	-7.569	-54.335	-40.245	12.682	-3,407.107	2,976.751	52.417	45.796	1.096	0.639	-191.301	0.008	5.295	5.127	0.643	53.308	1.106	52.697	98.712	45.
32	256.373	2.521	1.340	56.055	115.037	-8.894	-64.362	-37.723	13.513	-4,411.957	3,094.621	67.876	47.610	1.287	0.684	-187.210	0.015	5.302	5.133	0.000	0.000	0.000	0.000	0.000	0.
33	207.009	2.183	1.262	52.809	99.015	-7.198	-54.064	-39.873	12.522	-3,436.137	3,024.714	52.864	46.534	1.114	0.644	-188.381	0.021	5.308	5.140	0.650	54.068	1.127	53.236	100.191	46.
34	203.968	2.212	1.275	53.336	100.099	-7.044	-54.340	-39.845	12.534	-3,489.721	3,041.848	53.688	46.798	1.129	0.651	-187.372	0.027	5.314	5.146	0.653	54.431	1.139	53.504	100.886	46.
35	242.957	2.499	1.359	56.841	113.365	-8.206	-63.135	-35.760	13.363	-4,429.663	2,967.694	68.149	45.657	1.276	0.694	-181.558	0.034	5.321	5.152	0.000	0.000	0.000	0.000	0.000	0.
36	203.113	2.235	1.286	53.791	99.083	-4.998	-54.841	-38.895	12.653	-3,543.089	2,922.117	54.509	44.956	1.141	0.656	-184.855	0.046	5.333	5.165	0.659	55.224	1.151	53.976	99.816	44.
37	221.106	2.166	1.244	52.283	98.387	-6.892	-55.634	-40.286	12.562	-3,476.092	2,943.798	53.478	45.289	1.111	0.638	-192.015	0.002	5.006	5.384	0.641	54.348	1.123	52.545	99.468	45.
38	218.559	2.192	1.246	52.364	100.619	-8.064	-55.161	-40.744	12.737	-3,476.479	3,089.085	53.484	47.524	1.124	0.639	-190.836	0.009	5.012	5.390	0.642	54.492	1.136	52.614	101.658	47.
39	214.807	2.215	1.251	52.550	101.957	-8.406	-54.957	-41.146	12.640	-3,465.964	3,186.949	53.323	49.030	1.136	0.641	-189.014	0.015	5.019	5.397	0.645	54.300	1.147	52.835	103.012	49.
40	400.967	0.287	4.008	168.410	19.371	-5.951	-60.146	-35.134	8.071	-1,409.479	-158.371	21.684	2.436	0.147	2.055	-179.809	0.021	5.025	5.403	0.000	0.000	0.000	0.000	0.000	0.
41	205.661	2.206	1.270	53.352	100.656	-7.494	-54.877	-40.218	12.462	-3,524.898	3,043.034	54.229	46.816	1.131	0.651	-187.581	0.028	5.031	5.409	0.651	55.349	1.160	53.350	103.164	48.
42	201.320	2.181	1.270	53.367	98.953	-6.816	-53.935	-39.822	12.374	-3,466.274	2,990.441	53.327	46.007	1.118	0.651	-187.406	0.034	5.038	5.416	0.656	54.628	1.137	53.731	100.491	46.
43	203.570	2.212	1.283	53.923	99.003	-5.544	-54.404	-39.464	12.495	-3,508.662	2,951.328	53.979	45.405	1.134	0.658	-184.615	0.047	5.050	5.428	0.661	54.810	1.145	54.191	99.852	45.
44	540.932	13.803	36.490	1,543.660	490.295	98.290	-392.891	-195.421	5.824	-40,538.935	-8,665.191	623.676	133.311	7.126	18.838	-171.683	-0.006	4.723	5.647	0.000	0.000	0.000	0.000	0.000	0.

Introduction

- **To this:**



What is a Dataset

- Comprised of two parts:
  - Column names, referred to as "Features"
  - And data points (rows)
- Often stored as a ".csv" file
  - feature1, feature2, feature3 [first row called a header]
  - d1, d2, d3 [all other rows referred to by their index]

Front Ride Height	Rear Ride Height	Downforce	Chassis Angle	Chassis Heave
6.442514877	4.089697725	101.3770801	-1.8131	-0.1429
5.66649446	4.811475098	114.7736165	-0.9146	-0.1429
5.278356263	5.172482827	120.3391432	-0.4653	-0.1429
4.825570796	5.593619047	125.4008859	0.0588	-0.1429
4.372703189	6.014831666	123.9335348	0.583	-0.1429

Interpreting a Dataset

- What can you tell from this dataset?
  - How do values change with respect to other features
- What do you want to know?
  - Why are you collecting these features?

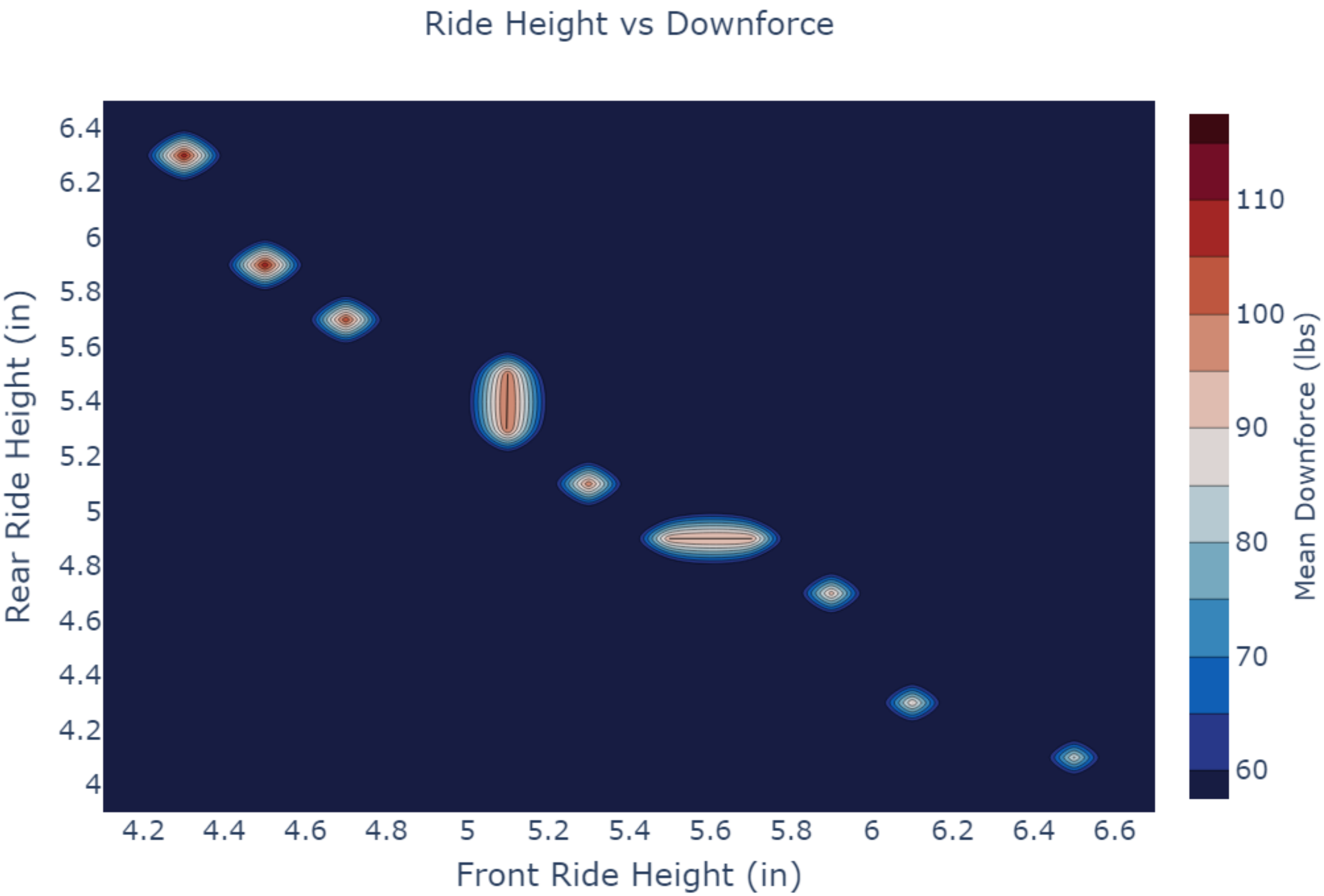
Front Ride Height	Rear Ride Height	Downforce	Chassis Angle	Chassis Heave
6.442514877	4.089697725	101.3770801	-1.8131	-0.1429
5.66649446	4.811475098	114.7736165	-0.9146	-0.1429
5.278356263	5.172482827	120.3391432	-0.4653	-0.1429
4.825570796	5.593619047	125.4008859	0.0588	-0.1429
4.372703189	6.014831666	123.9335348	0.583	-0.1429

Obtaining a Dataset / Testing Plans

- More data is always better
  - It may sound useless to collect but it is probably not
  - Wind direction may seem pointless, but we can derive a lot of cause and effect by such a simple measurement
  - Keep all of the data until it is certain that it is not needed
- Plan out ahead of time every piece of data you want to collect
  - In your testing day plans, record everything
  - For our purposes, film the entire time the car is driving
  - Record any driver feedback
  - Have a timer going the entire session to help correlate later

Understanding what you are trying to analyze

- What do you need to make this graph?
- What do you need to fix change from the dataset to do this?



Looking back at the original dataset

Looking only at the features we need.

- What do you notice? What does the data say about the car?
- Can we use this data right now?

Front Ride Height	Rear Ride Height	Raw Downforce Mean
6.421	4.068	0.000
6.427	4.074	0.000
6.433	4.081	85.007
6.440	4.087	85.315
6.446	4.093	85.832
6.452	4.100	86.770
6.465	4.112	87.592
6.138	4.331	0.000
6.144	4.337	0.000

Subset for easier viewing here.

Looking back at the original dataset

How can the car have 0 downforce?

- What went wrong?
- How do we fix this?

Applying your knowledge

- All these data points came from CFD simulations
- 2 main reasons behind this
  - The ride heights were out of bounds, meaning the car was too high/low to exist
  - Or the simulation failed from some other error
- Knowing this, is it reasonable to remove the rows with 0 df?

Intro to Data Cleaning

- Very few datasets ever start out usable
  - How do we clean a dataset? What does it mean to clean one?
- This is a very analytical process with a lot of trial and error
  - How to clean the data depends on domain knowledge from each sub team
  - Your cleaning process will probably be wrong the first time or even few times if the dataset is more complex
  - It is okay to be wrong here, just be sure to think about why it is wrong and what you can do better

Intro to Data Cleaning

- Basic Process

1. Convert any categorical features to numerical
   - Often hard to work with a word vs a number
2. Fill or remove any null or invalid values
   - Can you fill in the null values?
   - Or remove them
3. Are there redundant or useless points
   - Very situational
   - Only use if you know what you are doing
   - Example, the car is idling on track for 30 seconds before we start driving, this will mean the first 30 seconds can possibly be removed.
4. Advanced / Situational
   - Normalize the data
   - Do you need the data in a different form?
   - Unit conversions

Intro to Data Cleaning

- Types of features
  - Numerical
    - Any type of number for our purposes
  - Categorical
    - Wind direction, weather conditions
    - Driver
- Conversion Method (basic)
  - Ordinal Encoding
    - Assign each value a unique number
    - Wind directions
      - 1 = North
      - 2 = East
      - 3 = South
      - 4 = West
  - Many more but for our purposes for just analysis this is the easiest

Intro to Data Cleaning

- Types Null removal
  - Imputation
  - Removal

- Imputation
  - Fill in the null value with a value
  - Mean, median, mode, custom, 0 fill
  - Mean: fills in any null value in the feature with the mean value across that feature

- Removal
  - Remove the whole row

DALLAS FORMULA RACING

Intro to Data Cleaning

- **Normalization**
  - Converting the data to a value between [0,1] (most often 0,1)
  - Converts the data to be proportional and equal weight for all features
    - House prices in 100,000 will outweigh and negate a feature like number of bedrooms in a house
    - If both values are scaled between [0,1] they each have proportional weight
  - Each feature is normalized individually for this to work
  - May not use this much for our data

Next Steps

- We will cover the rest next lectures
  - Visualization
  - Analysis
  - Correlation / Validation
  - Advanced methods

Live Walk Through

Any questions before I start?