ChBE 312

Final Group Project:
It's α = 0.05 o'clock Somewhere

Group 7:
Antonio Hazboun,
Liz Schauer,
Jacqueline Vega

18 March 2019

Introduction

It seems like there is an endless amount of data available on the web, and finding a topic can be challenging. We first started by looking at data available on government websites, such as data.gov. After reading an article from the Wall Street Journal about Bud's Super Bowl ad potentially derailing the beer alliance, we found a graphic depicting the trends in US market share by alcohol type. It is visually apparent in Figure 1 below that throughout the last 20 years, the market share of wine has stayed relatively constant.

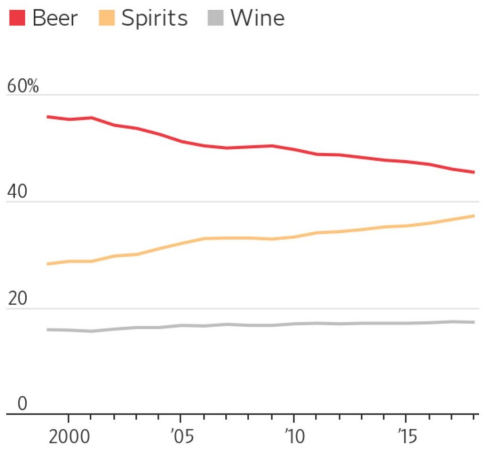

Figure 1. U.S. Market share by revenue. Source: WSJ.

We first questioned if there were different wine variety preferences over this period, or if these preferences stayed constant. Although we could not find information about wine preferences over time with enough data to analyze, we found a data set with over 130,000 entries on Kaggle.com (https://www.kaggle.com/zynicide/wine-reviews). The data set includes all wine reviews on winemag.com as of November 2017. With the information provided in this data set, we shifted the focus of our project to find the best value wines (the highest scoring wines for the cheapest cost). After doing some initial research and data analysis, we found that there are many factors that determine the score of a wine, including the taster that reviewed it, the country from which the wine originated, and the type of wine, among others. The tasters are all professional reviewers. It should be noted that this data set only includes scores between 80-100 points (it does not include values from 0-79) and scores are assigned integer values. Thus, we may not be able to draw conclusions about the distribution of all scores, as only the upper parts of the distribution are known.

Initial Investigation

As seen below in Figures 2, it is difficult to see trends when the data is broken down for only one category. Although price generally increases with increasing score, defining this relationship precisely is challenging.

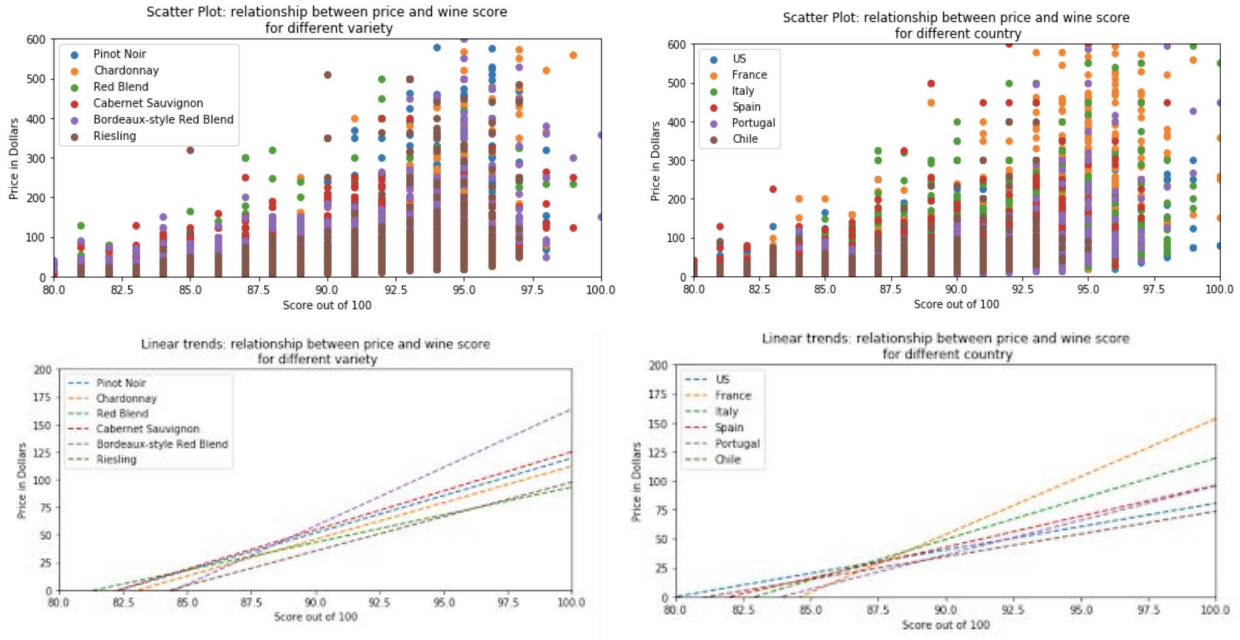

Figure 2. Left: Relationship between price and wine score for different varieties. Right: Relationship between price and wine score for different countries.

The r^2 correlation coefficients between scores and prices across all subcategories ranged from 0.27 to 0.55 with an average of $r^2 = 0.42$. This demonstrates that there is not a strong predictive relationship between price and score alone as the score cannot account for most of the variation in price. As such, other factors must be considered.

We then subdivided this problem further by investigating scores for a certain country or variety for a specific taster to see if we could extract more meaningful trends. However, we realized that the average scores for items in a category vary by taster. Thus, our project question became:

Do wine tasters exhibit bias for certain countries or varieties?

Methodology

In order to explore this question, we sorted the countries and varieties by number of reviews and did analysis on the top 10 in each category. We obtained the scores for each country and variety for a taster and determined the mean and standard deviation for each sample. These results were then plotted.

Results and Discussion

<u>Average Scores for varieties and countries for different tasters:</u>

The first taster we analyzed was Roger Voss. He had the most wine reviews of all tasters, with a total of 20,172 reviews. According to Voss's profile, he covers Bordeaux, Burgundy, Champagne, the Loire, southwest France, and Portugal. He has also written six books on wine and food. We did analysis on both the score given based on country and based on type of wine. These information is represented in the figure below:

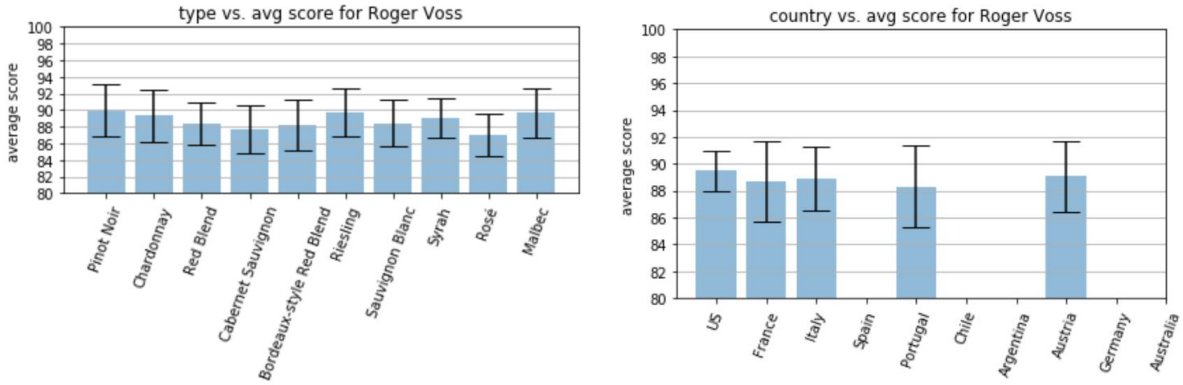

Figure 3. Left: Type of wine vs. average score for Voss. Right: Country vs. average score for Voss.

From Figure 3 above, it is visually apparent that his scores vary more for variety than country. Although the country scores are similar, there might be a bias against Portugal. Because Voss is an expert on Portuguese wines, he may be more reluctant to give a high score to wines from Portugal.

The second taster we analyzed was Joe Czerwinski because he appeared to have a bias against wine from the United States from initial analysis. He is from the United States; however, from his profile he is an expert on New Zealand, Australian, and southern France wines. Unfortunately, there are not enough wines from New Zealand for meaningful analysis. He reviewed a total of 5,010 wines in this data set, and his average score by type and country are shown below:

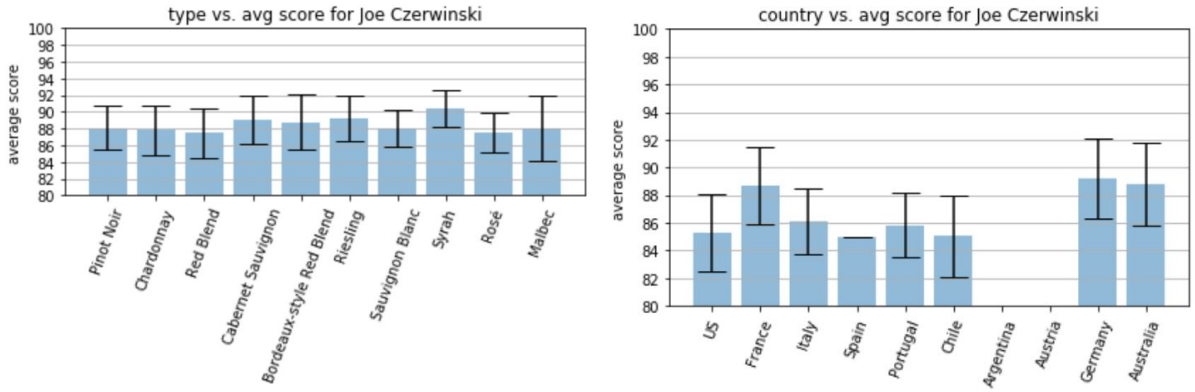

Figure 4. Left: Type of wine vs. average score for Czerwinski. Right: Country vs. average score for Czerwinski.

Although Voss appears to give lower scores to wines in his areas of expertise, it is clear from Figure 4 that Czerwinski rates wines in his specialty higher than those outside of his specialty.

The last taster we want to highlight is Michael Schachner. He has reviewed 14,944 wines in this data set and seems to have a bias against French wines on first inspection. He is from New York, so his origin may play a role in this bias. Schachner specializes in South American and Spanish wines. His average score by type and country are shown below:

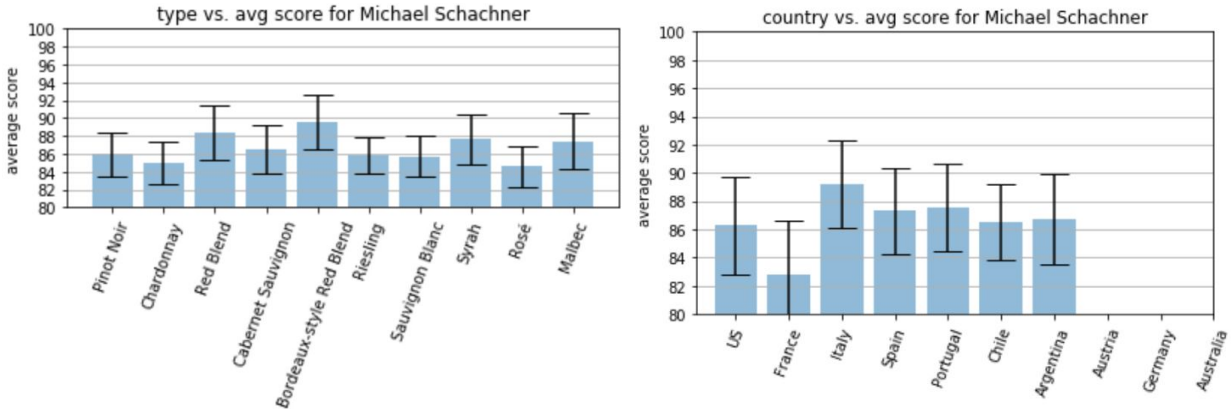

Figure 5. Left: Type of wine vs. average score for Schachner. Right: Country vs. average score for Schachner.

It is clear in Figure 5 above that Schachner's scores for varieties and countries vary widely. Although it appears that Schachner has a strong bias against French wines, this bias may not be statistically significant, as he only reviewed 20 French wines. To test this bias, Anova and Welch's T-Test were used, which are explained below.

Anova and Welch's T-Test Results:

The Anova Test and Welch's T-Test are used to test the hypothesis that two populations of data have the same mean. These tests take the relationship between the variation between populations (difference in means) and the variation *within* samples (variance as calculated by standard deviation). We first used the Anova test because it keeps the Type I error at 5%. However, it assumes that the samples have equal variance. Looking at the error bars of our bar charts, we can see from inspection that this is largely the case. However, just to make sure, we repeated the analysis using Welch's T-Test, which does not make this implicit assumption to see if the results agree with those of the Anova Tests. Welch's T-Test does have a high Type I error, so only magnitude should be considered when looking at its values.

The two populations used in both tests are:

1. The data for that particular subcategory (wine of a certain variety or country)
2. The data for all remaining subcategories

If the magnitude of the outputs of the Anova Test or Welch's T-Test are different, then there is reason to suspect that the taster might be biased for one type of wine or country over another. This difference is meaningful if the P-value is low (below 0.05 is the standard threshold). If the test results are large for all of the samples, then the ratings assigned by this taster are highly inconsistent across countries. The first sample analyzed was taster scores differences across wines from different countries, shown below:

Table 1. Summary of F and P values for taster scores across different countries for Anova test.

		US	France	Italy	Spain	Portugal	Chile	Argentina	Austria	Germany	Australia
Roger Voss	F	0.177	4.01	0.601	NaN	34.3	NaN	NaN	20.0	NaN	NaN
	P	0.674	0.0453	0.438	NaN	4.87e-09	NaN	NaN	7.67e-06	NaN	Nan
Michael Schachner	F	3.36	36.2	38.2	74.0	0.386	62.6	15.5	NaN	NaN	NaN
	P	0.0670	1.82e-09	6.41e-10	8.48e-18	0.53	2.63e-15	8.34e-05	NaN	NaN	NaN
Joe Czerwinski	F	122	0.613	61.0	1.46	15.5	32.5	NaN	NaN	14.0	5.36
	P	6.83e-28	0.43	7.24e-15	0.23	8.47e-05	1.27e-08	NaN	NaN	1.83e-04	0.021
Anna Lijima	F	298	NaN	NaN	NaN	NaN	NaN	NaN	NaN	429	NaN
	P	2.63e-65	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.15e-92	NaN

The NaN values indicate that there is no data for that subcategory by that taster. It is visually apparent that different tasters have different styles of biases.

- Tasters with a bias for or against one country:
    - The main example is Voss, who seems to show relatively small F values for wine from each country except for Portugal and Austria. From inspection of Figure 3, we notice that these countries have slightly lower means than the other countries from which he reviews. The P values for Portugal and Austria are are low. Thus, we can conclude that Voss is giving lower scores to wines from Portugal and Austria.
- Tasters with no clear trend or consistency in their ratings:
    - The main example is Schachner, who has large F values for almost all of the countries. The US and Portugal have low F values because their mean is close to the mean of remaining combined data. These large F values and small P values show that Schachner may not show bias towards one group.
    - This pattern is also observed for Lijima and Czerwinski, who both appear to not have bias with regards to countries.

After analyzing the Welch's T-Test values, we see agreement between the Welch's T-Test values and the Anova values (Appendix A). There is little consistency for Schachner and Czerwinski, although less inconsistency than was observed with the Anova test. We also see again that Portugal and Austria are the only countries for which Roger Voss is biased.

We then performed the same analysis for each taster and wines of different varieties, the table summarizing the F values for the Anova Tests is shown below:

Table 2. Summary of F and P values for taster scores across different varieties for Anova test.

		Pinot Noir	Chardonnay	Red Blend	Cabernet Sauvignon	Bordeaux-style Red Blend	Riesling	Sauvignon Blanc	Syrah	Rosé	Malbec
Roger Voss	F	220	98.0	2.20	3.33	50.5	41.9	3.46	0.731	409	43.0
	P	2.50e-49	5.00e-23	0.138	0.0682	1.25e-12	1.00e-10	0.0627	0.393	1.58e-89	5.81e-11
Joe Czerwinski	F	8.70	20.3	9.82	12.7	0.970	34.8	8.59	100.	22.3	0.139
	P	3.21e-03	6.72e-06	1.74e-03	3.71e-04	0.325	4.11e-09	3.40e-03	2.53e-23	2.47e-06	0.709
Anna Lijima	F	8.95	157	27.6	23.4	11.1	72.4	22.5	4.57	33.6	0.0107
	P	2.79e-03	3.04e-35	1.63e-07	1.40e-06	8.63e-04	2.23e-17	2.18e-06	3.27e-02	7.48e-09	0.917
Kerin O'Keefe	F	NaN	0.778	0.0637	0.223	NaN	0.0142	1.60	1.04e-02	9.06	NaN
	P	NaN	0.378	0.801	0.637	NaN	0.905	0.206	0.919	2.64e-03	NaN
Jim Gordon	F	137	20.4	24.0	5.33	1.16	5.70	37.6	46.6	13.1	0.607
	P	5.31e-31	6.43e-06	1.02e-06	0.0210	0.281	0.0170	9.93e-10	1.10e-11	3.08e-04	0.436

It is important to note the apparently limited relationship between inconsistency amongst scores for country of origin and inconsistency amongst scores for variety of wine. For example, Voss, who has one of the most consistent means across countries of origin, actually shows great inconsistency amongst his scores for different types of wine, and most of these values are statistically significant. Czerwinski, Lijima, and almost all other tasters have larger F values for their scores across wine varieties than across countries of origin. This is noteworthy seeing that the variety of wine influences how the wine is made and the country of origin impacts the grapes and land from which the raw materials are sourced. This suggests that the method of preparing wine has more of an impact on the opinions and biases of tasters.

One interesting insight is that O'Keafe and Gordon have the most equal distributions of samples across wine varieties. This statement is highly significant seeing that O'Keafe and Gordon only reviewed wines from Italy and the United States. This insight suggests either that wines more closely resemble each other when they are sourced from the same grapes and land, leading to less variation in score distribution or that dedicating one's career to reviewing wines from one place indicates more specialization and mastery, which could contribute to less bias and control over one's palate. It is visually apparent that O'Keafe is the most consistent and unbiased taster (although her P values are high), and Gordon has a statistically significant bias towards Pinot Noir.

After analyzing Welch's T-Test values, according to both the Anova and the Welch T-test, Kerin O'Keafe is the most consistent taster with no bias across wine varieties. However, Jim Gordon, who was previously the second most consistent when tested with Anova, is seen as being much less consistent and more biased across varieties when tested with Welch's T-Test. The corresponding P values demonstrate that Jim Gordon is in fact less consistent when assessed with the T-Test.

Conclusion

Overall, we were able to conclude that some tasters exhibit bias for certain countries or varieties. However, the bias varies greatly by taster. Some tasters, like Voss, display bias against wines in their country areas of expertise; however, others, like Czerwinski, do not appear to have a clear trend or consistency in their ratings with regards to countries. There is a limited relationship between inconsistency amongst scores for country of origin and inconsistency amongst scores for variety of wine, as Voss shows less consistency among varieties than countries. Gordon is one of the least biased tasters with regards to variety; however, it appears that he has a bias towards Pinot Noir.

For future analysis, we would want to investigate the bias between raters for a specific country or a specific variety. One report showed that red wines are rated higher than white wines (https://www.winecurmudgeon.com/expert-scores-red-wine-bias/), so it would be interesting to see if this data also supports that claim. Another factor to consider is the year the wine was made, as the grapes could have been grown in different conditions for the same area. The temporal element could be an interesting dimension to see if certain years are better or worse than others.

Appendix A: Welch's T-Test results

Table 3. Welch's T-Test F values for different countries.

	US	France	Italy	Spain	Portugal	Chile	Argentina	Austria	Germany	Australia
Roger Voss	0.420857	2.00171	0.775553	NaN	-5.85372	NaN	NaN	4.4753	NaN	NaN
Michael Schachner	-1.8319	-6.01693	6.18412	8.60047	0.621681	-7.91365	-3.93536	NaN	NaN	NaN
Kerin O'Keefe	NaN	NaN	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Virginie Boone	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Paul Gregutt	0.0448982	0.204572	-1.29389	-2.35617	NaN	NaN	NaN	NaN	NaN	NaN
Matt Kettmann	0.00360706	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Joe Czerwinski	-11.0344	0.783163	-7.81162	NaN	-3.93492	-5.70334	NaN	NaN	3.74425	2.31484
Sean P. Sullivan	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Anna Lee C. Iijima	-17.258	NaN	NaN	NaN	NaN	NaN	NaN	NaN	20.709	NaN
Jim Gordon	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Table 4. Welch's T-Test P values for different countries.

	US	France	Italy	Spain	Portugal	Chile	Argentina	Austria	Germany	Australia
Roger Voss	0.673864	0.0453238	0.438022	NaN	4.86696e-09	NaN	NaN	7.67054e-06	NaN	NaN
Michael Schachner	0.0669871	1.81969e-09	6.40938e-10	8.48155e-18	0.534161	2.63603e-15	8.33752e-05	NaN	NaN	NaN
Kerin O'Keefe	NaN	NaN	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Virginie Boone	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Paul Gregutt	0.964189	0.837911	0.195737	0.0184852	NaN	NaN	NaN	NaN	NaN	NaN
Matt Kettmann	0.997122	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Joe Czerwinski	6.82763e-28	0.433569	7.24334e-15	NaN	8.47346e-05	1.26553e-08	NaN	NaN	0.00018342	0.0206568
Sean P. Sullivan	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Anna Lee C. Iijima	2.62732e-65	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.14592e-92	NaN
Jim Gordon	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Table 5. Welch's T-Test F values for different varieties.

	Pinot Noir	Chardonnay	Red Blend	Cabernet Sauvignon	Bordeaux-style Red Blend	Riesling	Sauvignon Blanc	Syrah	Rosé	Malbec
Roger Voss	14.8327	9.90103	-1.48288	-1.82411	-7.10692	6.47326	-1.86138	0.854865	-20.2335	6.55535
Michael Schachner	-6.24226	-17.1454	18.0917	-4.16459	10.5676	-1.62387	-10.0009	5.09836	-11.2141	7.01811
Kerin O'Keefe	NaN	-0.882291	0.252492	0.472629	NaN	0.119081	-1.26367	0.101857	-3.00941	NaN
Virginie Boone	8.36502	0.979872	-8.91828	0.312975	0.855685	-1.39794	-9.57279	0.399836	-3.92336	-2.2071
Paul Gregutt	2.89936	-0.581866	-6.80203	-0.212858	6.20949	0.527055	-3.68378	1.68178	-6.34281	-3.14999
Matt Kettmann	10.0345	-3.29135	-2.39849	-6.19834	1.65848	-4.63507	-10.0948	5.02565	-8.61634	-2.94216
Joe Czerwinski	-2.94913	-4.50958	-3.13448	3.5637	0.985034	5.89513	-2.93127	10.0301	-4.71933	-0.373244
Sean P. Sullivan	-4.22835	-7.23054	-3.97933	1.5775	9.16524	-8.20616	-9.07312	11.8576	-6.96333	-2.04414
Anna Lee C. Iijima	-2.99238	-12.5312	-5.24922	-4.83425	-3.33491	8.5112	-4.74443	-2.13727	-5.79517	-0.103627
Jim Gordon	11.7124	-4.52033	-4.89762	-2.30898	1.07734	-2.38826	-6.13302	6.82414	-3.61323	0.779321

Table 6. Welch's T-Test P values for different varieties.

	Pinot Noir	Chardonnay	Red Blend	Cabernet Sauvignon	Bordeaux-style Red Blend	Riesling	Sauvignon Blanc	Syrah	Rosé	Malbec
Roger Voss	2.50115e-49	5.0048e-23	0.138136	0.0681648	1.2462e-12	1.00123e-10	0.0627161	0.392646	1.58114e-89	5.80602e-11
Michael Schachner	4.53575e-10	8.97739e-65	7.24065e-72	3.14913e-05	6.39824e-26	0.104446	2.05956e-23	3.50737e-07	5.86303e-29	2.4113e-12
Kerin O'Keefe	NaN	0.377698	0.800671	0.636517	NaN	0.90522	0.206462	0.918878	0.00264273	NaN
Virginie Boone	6.936e-17	0.327178	5.9115e-19	0.754307	0.392201	0.162175	1.38644e-21	0.689289	8.81599e-05	0.0273398
Paul Gregutt	0.00374793	0.560674	1.1137e-11	0.831443	5.61957e-10	0.598172	0.000231576	0.0926541	2.40082e-10	0.0016398
Matt Kettmann	1.65611e-23	0.00100366	0.0165016	6.18405e-10	0.0972909	3.67072e-06	1.02932e-23	5.19632e-07	9.44991e-18	0.00327636
Joe Czerwinski	0.00320761	6.71677e-06	0.00173822	0.000371044	0.324687	4.10666e-09	0.00339806	2.52624e-23	2.4697e-06	0.708994
Sean P. Sullivan	2.41645e-05	5.82532e-13	7.04186e-05	0.114759	7.90935e-20	3.15582e-16	1.90185e-19	6.75869e-32	3.96615e-12	0.0410148
Anna Lee C. Iijima	0.00278871	3.03645e-35	1.62871e-07	1.39999e-06	0.000863105	2.22614e-17	2.18389e-06	0.0326541	7.48253e-09	0.917472
Jim Gordon	5.30956e-31	6.43036e-06	1.02497e-06	0.0210179	0.281435	0.0170049	9.93447e-10	1.10376e-11	0.00030836	0.435867