11/1 Meeting Report

Preliminary Information:

Standard Error - The standard deviation of the model's sampling distribution. This represents a measure quantifying the amount of variation within our data set. A lower standard deviation means that the points of our data are generally closer to the mean of the set, while higher values indicate that the points of our data are spread out over wider ranges. Thus, a lower standard deviation points towards a stronger trend in our data.

R-squared - Known as the coefficient of determination, represents the proportion of the variance in the dependent variable that is predictable from the independent variable, i.e., it shows the strength of correlation between variables in a linear fashion. R-squared is measured on an interval of [0, 1], where a value closer to 0 indicates a weaker correlation, and a value closer to 1 indicates a stronger correlation. Generally speaking, values between .3 and .5 show the existence of a correlation, with anything higher showing a strong correlation, and anything below denoting a very weak correlation that may simply be due to chance. In the context of this project, R-squared represents the reliability of an identified trend line (Mx + b) found within the data of one of the facial recognition softwares in determining the estimated age of a politician based on their actual age.

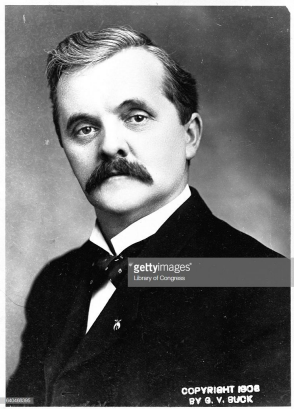
Trend Analysis for Amazon Estimation in Years

	CurrGov	2014Gov	CurrRep	CurrSen	'60Sen	'40Sen	'20Sen
Sample Size	130	78	95	176	91	153	130
Average Err	0.19	-0.40	-0.41	-0.47	3.6	4.97	5.01
Standard Err.	8.71656	10.667	10.7486	9.79416	9.14635	9.7747	10.8489
R-squared	0.356872	0.261537	0.143244	0.295389	0.308596	0.287987	0.141784

Trend Analysis for Microsoft Estimation in Years

	CurrGov	2014Gov	CurrRep	CurrSen	'60Sen	'40Sen	'20Sen
Sample Size	147	94	103	209	118	217	157
Average Err	-4.0748	-6.2553	-5.0097	-5.7512	-2.9915	-1.5899	-0.4395
Standard Err.	5.45872	8.13331	6.88184	6.19593	5.86148	6.29962	5.77181
R-squared	0.509295	0.235714	0.459752	0.579548	0.499773	0.533442	0.442331

Possible Reasons for Error and Variance:



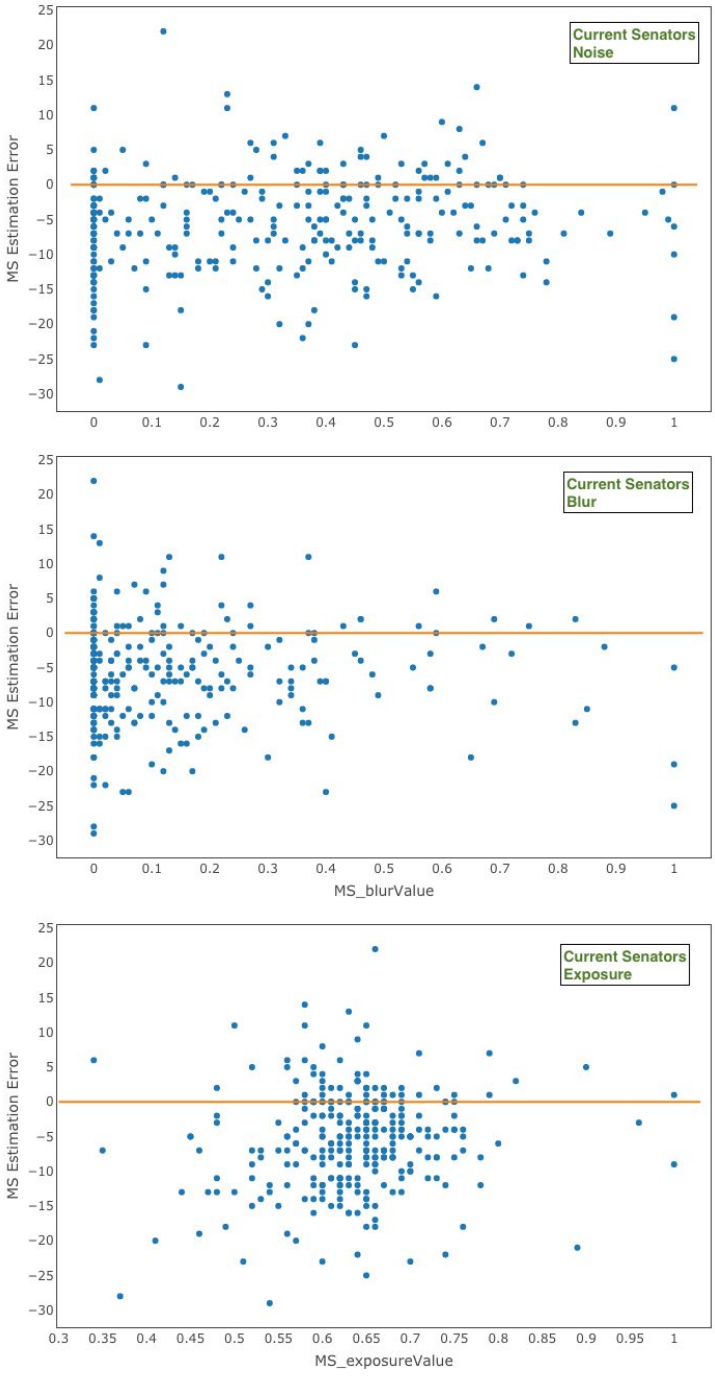
- While primarily an issue for images up through the ‘60s, washing-out of images causes age estimations for both Amazon and Microsoft to be skewed towards negative errors due to facial features being covered up by whiteness, causing skin to appear smoother and ‘younger’ than it should be. This ultimately causes features to not be picked up by the software, leading to underestimation becoming larger based on severity of the image being washed-out. The image to the right is an ideal example of how much estimation can be skewed by being washed-out as it is the worst example of an image being washed-out in our collection. While actually being 47 in the photo, Amazon estimated Senator Norris to actually be 12.5. However, this issue doesn’t solely cause estimation to skew negative, as it tends to also cause overestimation, but notably less frequently. Due to this issue caused by older technology, we believe this to be a primary source of R-squared values getting notably lower as the period goes further back.

- Shadows and glasses tend to be notable sources of overestimation in age for both softwares across all periods covered. The issue of shadows tend to come in two forms: Shadows from objects around the person being cast upon the subject’s face and poor lighting causing facial features to cast minor shadows that accentuate said features, causing the subject to appear older than they really are. The latter tends to be notably more common than the former since poor lighting is an issue that is hard to work around, even in situations where the photographer has full control over lighting and subject. In the issue of glasses, we believe glasses tend to be picked up by the software as shadows accentuating features, leading to the subject being estimated as older than they really are.

- Uneven lighting, i.e. one side of the face being darker than the other due to the primary light source not being placed in front of the subject, has a notable effect upon error in estimation in both directions for both softwares. As this seems to be the biggest issue among all periods, we believe this to be a notable source of variation, causing R-squared values to drop by a significant margin.

- When a subject’s face is turned just enough such that one side of their nose is no longer visible, estimations for both softwares tends to go into notably positive errors, generally at least ten years more than the actual age. Severity of error increased as the head was turned further away from the camera. While images used were chosen such that this would not be an issue, some were chosen such that there would be a larger sample size. As such, this may have caused a minor upward skew in average estimation across all periods. Significance upon set data needs to be tested.

- While believed to be a source of significant error prior to analyzing the Python data, blur, exposure, and noise had no significant effect on error for either Amazon or Microsoft in any period. Almost all R-squared values for noise, blur, and exposure for all datasets were under 0.01, with very few going above.
  (Attached below as an example are current senators' noise, blur and exposure values plotted against the microsoft estimation errors respectively. Their respective R-squared values for their trend lines were 0.0317, 0.00660, and 0.0251.)



Analysis:

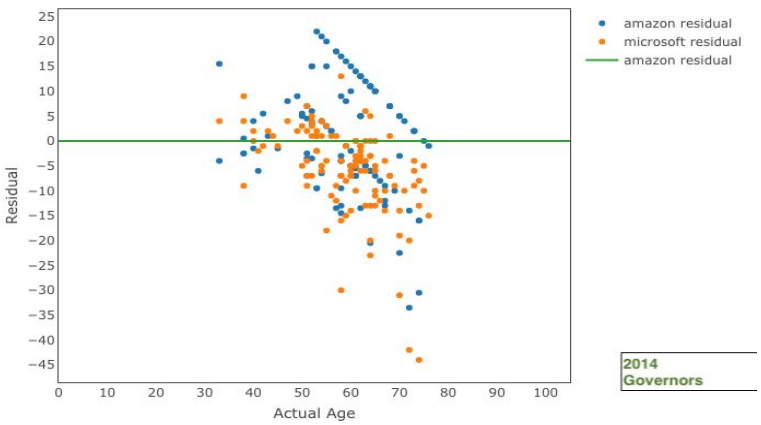
- For all current data sets (minus 2014Gov), Microsoft had consistently more severe errors with higher R-squared values ranging from 1.5 - 3x higher than that of Amazon's alongside having lower standard error. With this in mind, Microsoft's software is much more reliable in correlating a trend in rates of aging of modern politicians than Amazon's software.

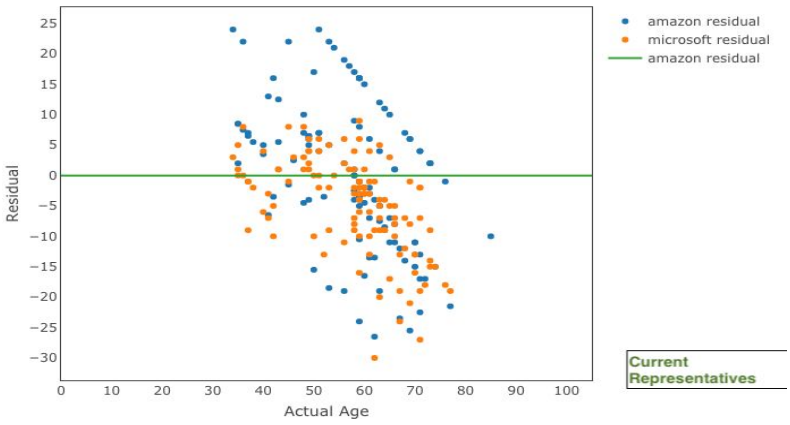
- For all older data sets ('60s and earlier), Microsoft had consistently less severe errors with higher R-squared values ranging from 1.5 - 3x higher than that of Amazon's alongside having lower standard error. With this in mind, Microsoft's software is much more reliable in correlating a trend in rates of aging of politicians in periods with less reliable photo technology than Amazon's software.

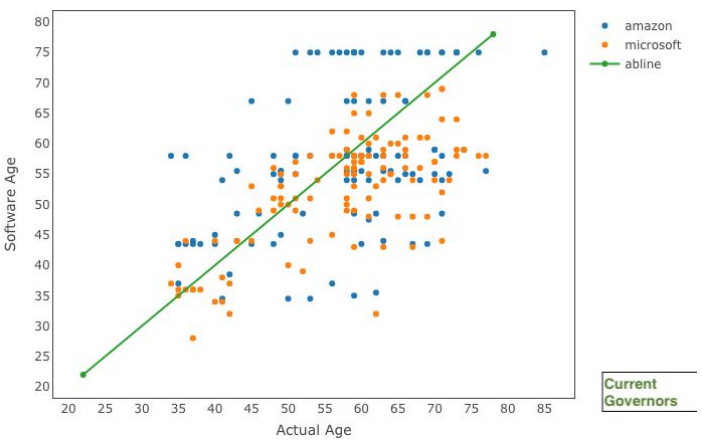
- Thus, Microsoft is a much more reliable software than Amazon when it comes to analyzing rates of aging in politicians across all periods due to consistently higher R-squared values.

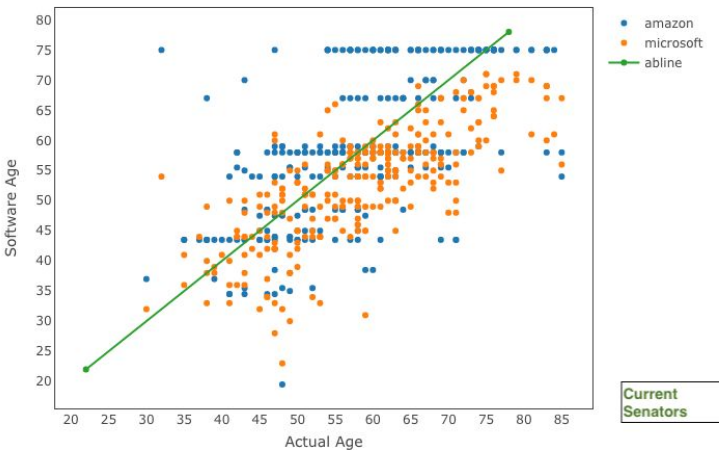
- As such, there seems to be a correlation that the rate of aging in politicians has actually decreased as time has passed, which is notable since the average of politicians in all datasets have been between 57 and 60.

- Both Amazon and Microsoft (as shown below) are more likely to overestimate the actual age before around 60, and more likely to underestimate the actual age after. They give a negatively sloped downward trendline in these residual plots.









- Amazon tends to assign the same age multiple times across all groups of politicians tested, giving several horizontal lines along 43.5, 58, 67, and 75.

- When Amazon error and Microsoft error are plotted against each other for any given dataset, the corresponding R-squared value of the trend line is never higher than 0.3, with 0.3 being an upper bound but not the supremum of the R-squared values. With all values being sub-0.3, it is very fair to say that the softwares were designed differently and do not process images in the same way at all. We believe that this is partially due to Amazon's aforementioned tendency to give estimates upon a select few ages rather than giving each photo a more 'unique' estimation of its own, similar to Microsoft's method of estimation.

- For current photos where both softwares had an absolute error less than three, not only did subjects look the age they actually were, but the photos were taken under ideal conditions in which subjects were looking head-on into the camera, had no shadows on their face, had no features accentuated by poor lighting, had no bright spots on their faces, were almost always intentionally taken portraits rather than action shots, and almost never had issues with image artifacting. We believe that this observation is not simply beneficial towards understanding the necessity of ideal lighting in a photo such that someone looks there age, but more so beneficial towards understanding just how heavily estimated age can be skewed by poor lighting.

Notes:

- Tableau does not plot multiple points of data sharing the same coordinates (i.e., more than one point with the same actual age and estimation), and thus, may not consider such repetitions in this data. Will have to check for accuracy manually via coding.
    - Additionally, the sample size on the charts above does *not* match the sample size on the Excel sheets due to Tableau not counting for repetitions. As such, the sample size given by Tableau is used above since it is safer to assume that we are using the smaller sample size with greater inaccuracy in the given data than to assume we're using a larger sample size with greater accuracy.