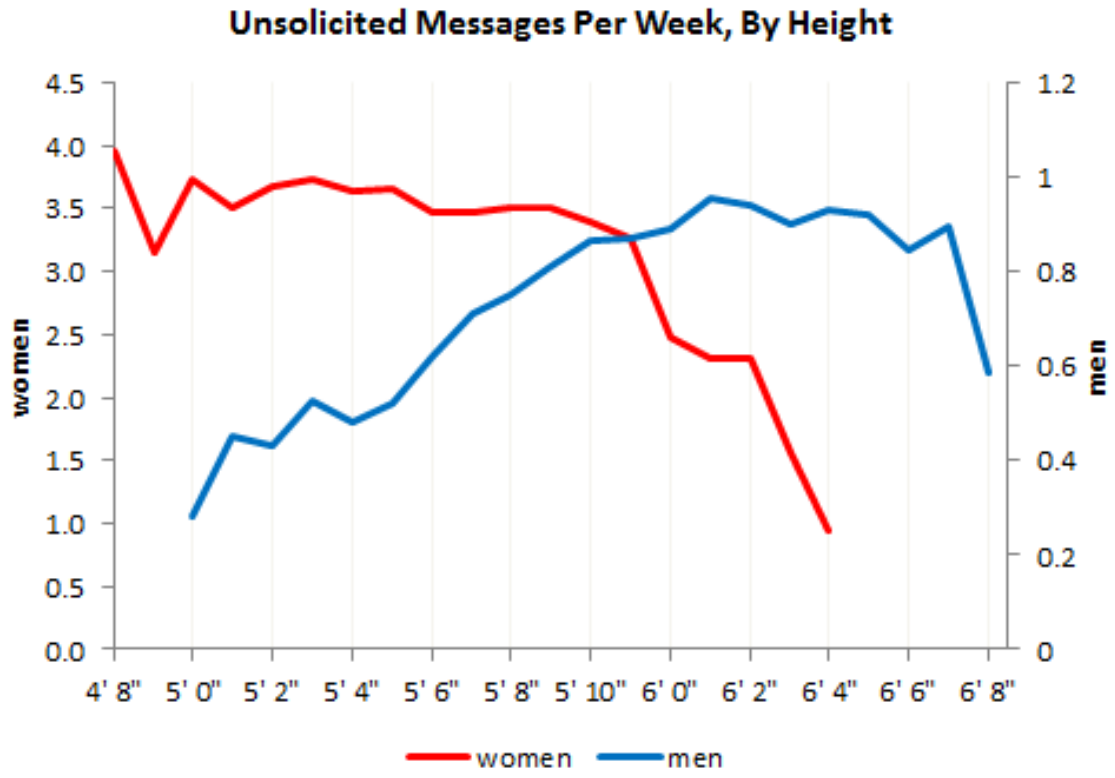
Perspectives on Computational Research

HW 3- Part 1: Visualization Critique

Given that I am working on the OkCupid dataset on dating behaviour, and that the height of male users forms a key component of one of my hypotheses, I have selected visualization for this exercise from the OkCupid blog (OkCupid, 2010). The topic too, fittingly, involves the relationship between height of a user and the number of unsolicited messages they receive, with results split by gender.

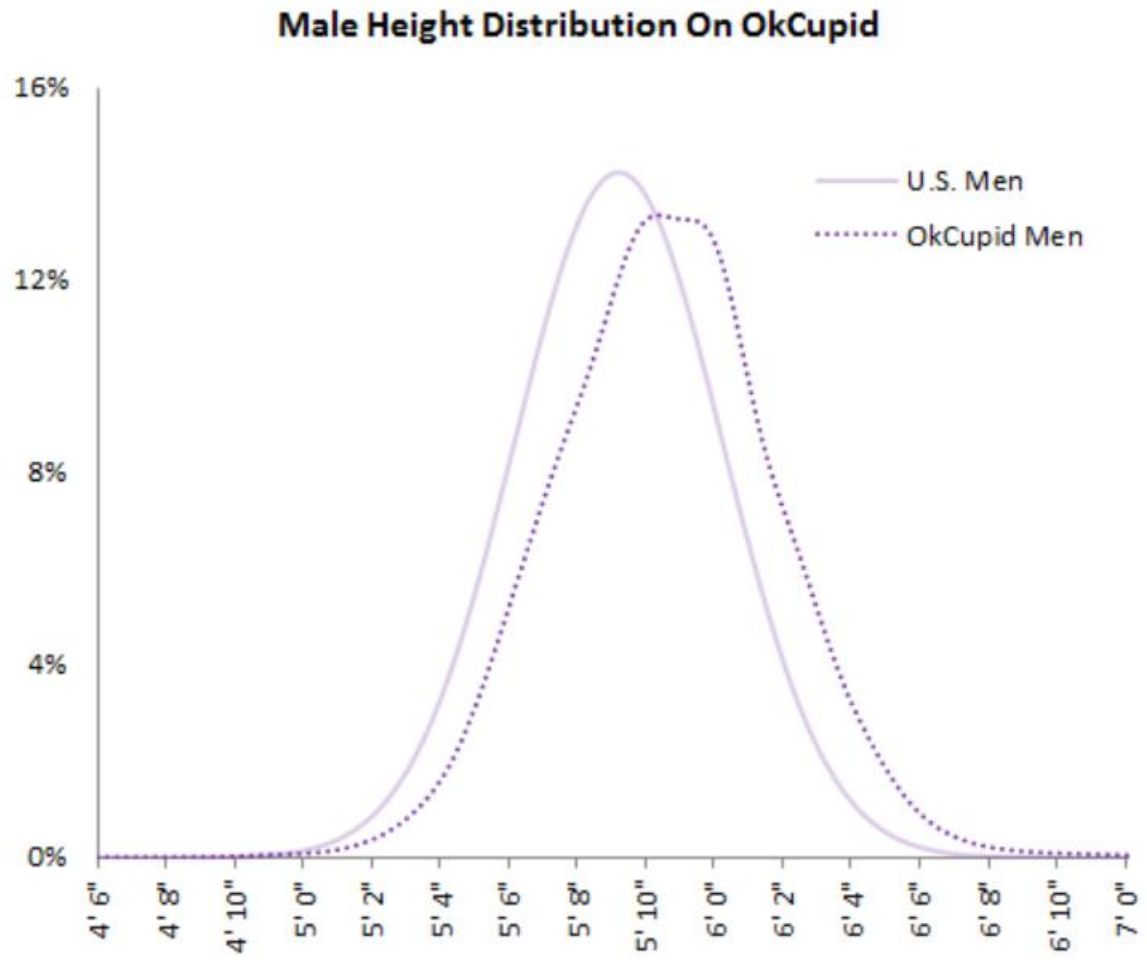


Criterion 1: Is it Truthful?

At first, the line graph does appear to speak about the relationship between height and number of messages. However, on closer inspection, it becomes clear many of the design and data choices may well be obscuring rather than illuminating facts.

The first consideration is the veracity of the reported data itself. As I am currently exploring in this research project, both male and female users have

been observed to overstate their heights. This tendency was unearthed by OkCupid's own data science team (on the very same blog post) by comparing the distributions of adults' heights across the US from the US census.



The shape of the distributions match, but their location does not. The entire distribution may be seen as 'shifted' forward by 2 inches. This would suggest that both men and women over-represent their heights. This is particularly the case around the benchmark of 6 feet. This may be seen more as a cultural benchmark for what qualifies as a 'tall' man. Women do this do, but to a lesser extent. Previous research also confirms this supposed preference for taller men among heterosexual women. So we cannot be sure of the accuracy of the data even to begin with. Ironically, the author does not reference their own finding in the same blog post.

This visualization also seems to operate on the assumption that both men and women's desire in the opposite sex can be measured using the same metric-

number of unsolicited messages a user receives. There may be far more complex mechanisms at play. For example, specific dating sites may offer the service of revealing which users have visited their profile. Others may provide features like 'wink' or 'wave'- which like their real world counterparts hint at their interest but still require the man to send a message. Visits, winks, and other metrics are not available here.

Furthermore, this may represent another case of correlation, not causality. Tall men may be used to special treatment, and hence exude confidence in their photos. It is that confidence- as observed in profile photos- and not the number visible under the 'height' question, that leads to more responses from prospective female romantic partners (Brand, 2012).

Criterion 2: Is if Functional?

For the most part, the chart does succeed in telling its story. However, it suffers from two issues.

Firstly, the chart could have been labelled more clearly, which leads to some confusion. For example, there is no mention of the measurement on either of the axes- neither the number of inches (for the x axis), nor the average weekly number of emails (on the y axis). The reader is left to infer this solely from the title.

Secondly, this y-axis variable becomes difficult to interpret. Of course, we are working with averages. Hence, the decimal figures of messages are acceptable). However, the chart would suggest that a man of 5 feet, 11 inches receives 5.7 messages. The findings become difficult to express in lay terms or any logically explainable manner.

Thirdly, the presence of two scales (on the right and left edges of the chart) for the same metric- the average number of messages- only complicates matters further. There appear to be distinct variables moving in opposite directions along the horizontal axis. In reality, they both move from left to right.

This design choice appears to have been intended to control for the fact that men women receive fewer unsolicited messages. This appears to be a trend driven by cultural, social and potentially even biological imperatives (on

‘making the first move’) (Lin & Lindqvist, 2013). Without this, the entire line for men would be limited between 0 and 1, as the maximum number of messages received for men appears to lie somewhere close to 7.

At this juncture, this quote comes from the same blogpost:

“a 5' 4" woman gets 60 more contacts each year than a six-footer”

With the adjustment of axes, the message is lost entirely- and calls truthfulness into question again. A previous paragraph states that women are desired at the height most at the height of about 5’4”, while men are desired most at the six foot mark. In spite of this, the most desired man receives considerably less messages than the most desired woman. Had this been the intended message, the men’s line should not have been scaled. After the designer’s scaling adjustment, the two lines seem to travel across a similar vertical range, and the entire disparity on receiving messages is lost.

Criterion 3: Is it Insightful?

It appears that the underlying meaning is supposed to emerge spontaneously rather through careful observation. The audience appears to be general, rather than academic. In that sense, the height metrics are easy to follow.

We see that the number of messages received flatlines across a range of heights before beginning to fall steeply at 5’10”. For men, the relationship between messages and height for appears positive and rising up till the six foot mark, where it plateaus and then falls again at the 6’2” mark. Thus a woman’s height appears to have limited importance until it crosses a specific threshold, after which the number of messages falls drastically. For men, height seems to hold importance throughout (both positive and negative) and only plateaus briefly.

To the author’s credit, this insight emerges clearly from the data.

Criterion 4: Is it Beautiful?

No, the visual is not particularly eye-catching. Given the simplicity of the line graph, I don’t think it held as much potential on this front.

In terms of Edward Tufte's Data-Ink Ratio, I believe the use of gridlines has been adequate. The vertical ones at different heights help keep track of the trajectory of the heights. That being said, it is simple and stark- not beautiful per se, but not in any way unsightly either.

Criterion 5: Is it Enlightening?

Yes, the visualization is enlightening. It does seem to confirm many popular beliefs that men are considered more attractive as they get taller. However, the limited relevance of height for women's attractiveness right until they hit the 5'10" mark seems surprising. Further evidence of this surprise comes from the earlier finding that women seem to overstate their height by 2 inches. From the results of these visualizations, it seems like female users would be better off leaving their original heights unadjusted.

REFERENCES

- Brand, R. J., Bonatsos, A., D'Orazio, R., & DeShong, H. (2012). What is beautiful is good, even online: Correlations between photo attractiveness and text attractiveness in men's online dating profiles. *Computers in Human Behavior*, *28*(1), 166-170.
- Lin, K. H., & Lundquist, J. (2013). Mate selection in cyberspace: The intersection of race, gender, and education. *American Journal of Sociology*, *119*(1), 183-215.
- OkCupid, 2010. 'The Big Lies People Tell In Online Dating'. Retrieved from: https://theblog.okcupid.com/the-big-lies-people-tell-in-online-dating-a9e3990d6ae2