Data Science with Python
Ticket Out the Door
Set 0: Visualizing Data

Name _____ Period _____

Skill 0.1 Exercise 1
What is data literacy?
Why is data literacy important?

Skill 0.2 Exercise 1
Provide another example of Garbage in, garbage out.

Skill 0.3 Exercise 1
Neural Machine Translation (NMT) is trained on example text that exists in the world. Consider the Google Translate that was constructed using NMT.

What is the bias in this translation.
How might Google Engineers modify the translation to remove bias?

Data Science with Python
Ticket Out the Door
Set 0: Visualizing Data

Name _____ Period _____

Skill 0.4 Exercise 1

In 2014, Amazon experimented with using software to screen job applicants. The screening software was trained on a decade of résumés that had been previously rated by employees as part of the hiring process. Below shows the distribution of male and female employees at Amazon in 2014.

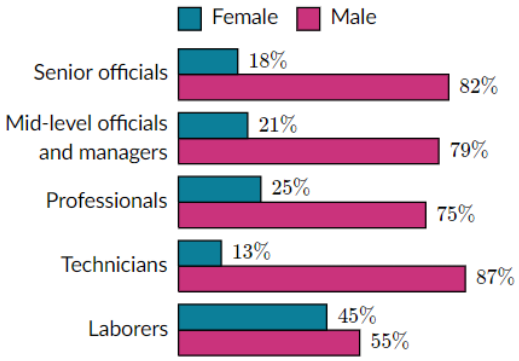


Chart source: Seattle Times

Is it statistically possible that the above results were by chance?

How might Amazon better train their software model to avoid "reckless regard" for the rights of female candidates?

Data Science with Python
Ticket Out the Door
Set 0: Visualizing Data

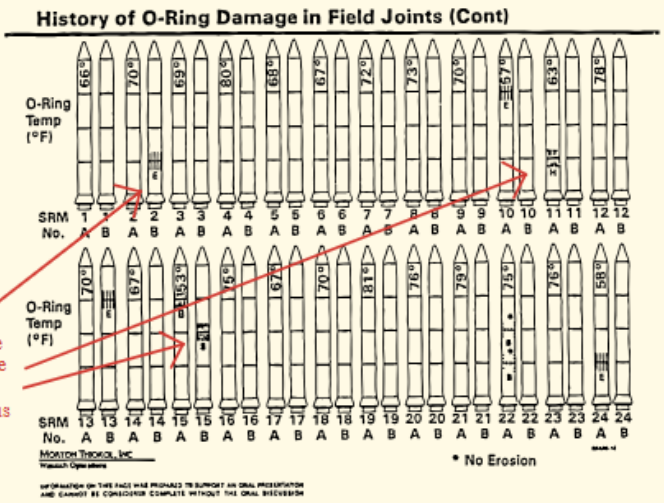
Name _____ Period _____

Skill 0.5 Exercise 1

On January 28, 1986 the Challenger space shuttle carried seven US astronauts who were supposed to deploy a satellite and study Halley's Comet while they were in orbit. Less than two minutes after takeoff, however, the shuttle exploded, killing all seven crew members. The explosion was caused by a failure of two O-rings: small rubber rings that helped create an airtight seal between the space shuttle and its launch fuel supply. Before the launch, engineers were concerned about how the low-temperature forecast would affect the O-rings' ability to make a proper seal. The engineers made their arguments in favor of postponing the launch using, in part, a series of data visualizations that showed launch success rates at various temperatures. Tragically, their arguments did not prevent the launch from proceeding.

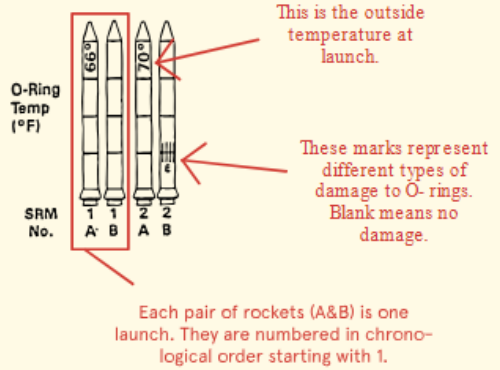
Below are the visualizations.

ORIGINAL: Rockets arranged by launch date



O-ring damage is spread in the visualization with no obvious pattern

How to read the charts:



How could the engineers have improved their visualization, and potentially prevented this tragedy?

Data Science with Python
Ticket Out the Door
Set 0: Visualizing Data

Name _____ Period _____