Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

U.S. DEPARTMENT OF ENERGY

Telling the Story: Gender Discrepancies in Engineering and Computer Science

APPROACH

Our approach was to leverage a relatively new technique called "storyline visualization" to convey patterns and trends in the dataset. In the past, storyline visualization techniques have been used to communicate interactions between individuals over time, which, for example, can come from scene colocation in a movie or novel. In our solution, we have used storylines in a manner that would typically be an application of "parallel sets." However, we chose not to use parallel sets; while parallel sets are good for conveying categorical information, it is difficult to trace entities through the visualization, which is useful to understand data that is changing over time.

We use storylines to convey what we call "common paths" in the visualization. These are combinations of particular attributes that many people tend to exhibit. For example, a fairly common path was a male who scored in the highest bracket on the ACT, declared and majored in engineering their first and second years. Our philosophy was to use storyline visualizations to understand the most common paths, which means that uncommon, or more unique individuals are not represented visually. In other words, this technique focuses on the major trends at the expense of ignoring outliers.

In these visualizations, the horizontal axis encodes different variables (i.e., columns from the dataset such as gender and ACT score). The vertical axis does not directly encode any information. Instead, it is used to group storylines together when they share a common attribute for a particular column. This allows us to more effectively see the relationships between the common paths of interest in, and the flow between groups over time. Furthermore, the vertical arrangement of storylines is carefully adjusted to make the visualizations easier to understand. Our custom multi-stage optimization algorithm is used to reduce effects that detract from the visualization's readability. The algorithm reduces line crossings, line wiggles, and unnecessary whitespace, which is a non-trivial combinatorial optimization problem.

This research is supported by the

Analysis in Motion Initiative

Technology advances have enabled the continuous capture and streaming of data from an increasingly diverse set of sources, from scientific instruments, to the web, to consumer electronics. Timely discovery, evaluation, and response to key insights from these streams require new computing capabilities and a rebalancing of human and computer effort. AIM is developing analysis and interaction technologies to enable humans to respond more quickly and accurately to data streams reflecting changing events.

Learn more at aim.pnnl.gov

CONTACT INFORMATION

Dustin Arendt, Ph.D.
Senior Visualization Scientist
dustin.arendt@pnnl.gov
509-371-6902

November 2015
PNNL-SA-112971

In collaboration with **Yanina Levitskaia**

UNIVERSITY of WASHINGTON

Telling The Story

Gender Discrepancies in Engineering and Computer Science

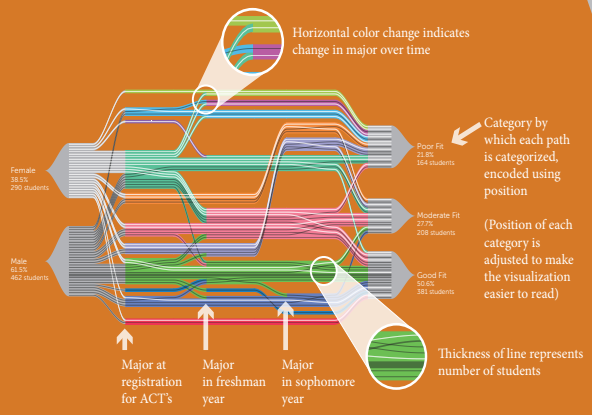
Data we are showing:

First, we looked at the relationships between score, major, and gender.

Then we took a closer look at the top two score brackets to see what paths men and women are taking and how they differ.

Last, we looked to see the outcome of students in computer science and engineering.

Reading the Visualizations



Overview of scores, majors, and genders

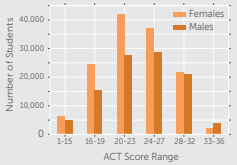
77,584 students took the top 35 most common paths.

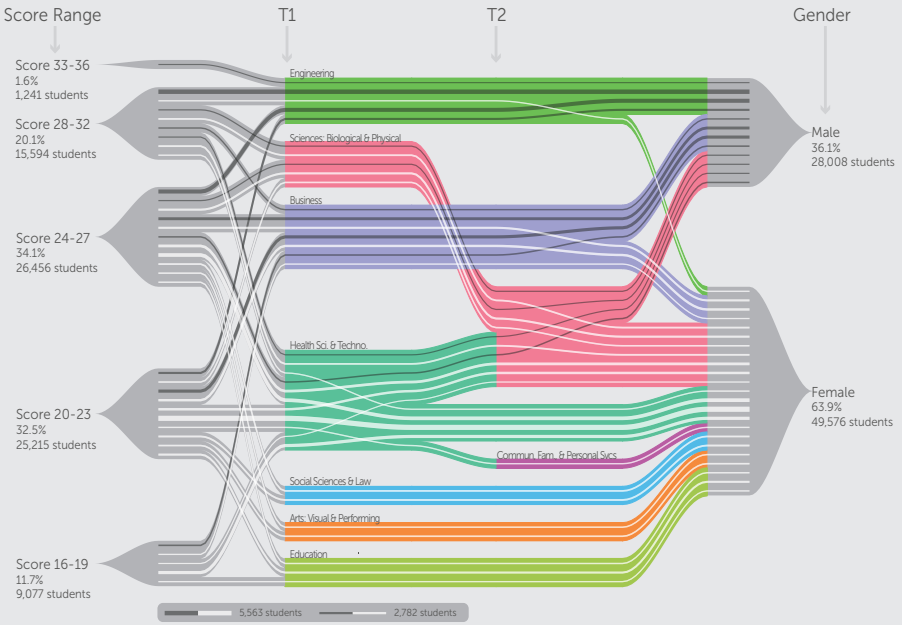
In the top most common paths taken, we see fewer women than men in the upper score ranges, especially in the 33 to 36 range. We also see that of those students who went into engineering, only a small fraction are women.

Could it be explained because the data has fewer women? Based on the graph below, there are actually more women than men in this data set, and many women did score in the 33-36 range.

Proportion of females to males within all data

Number of students within each score range, by gender

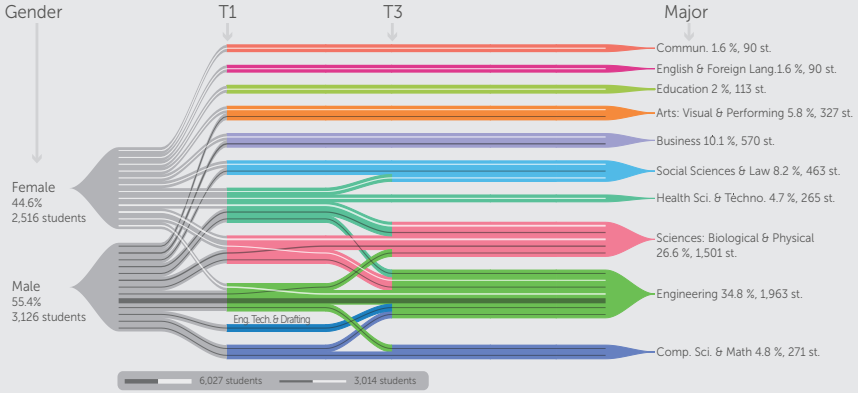




Students within the top two ACT score ranges (score of 28+)

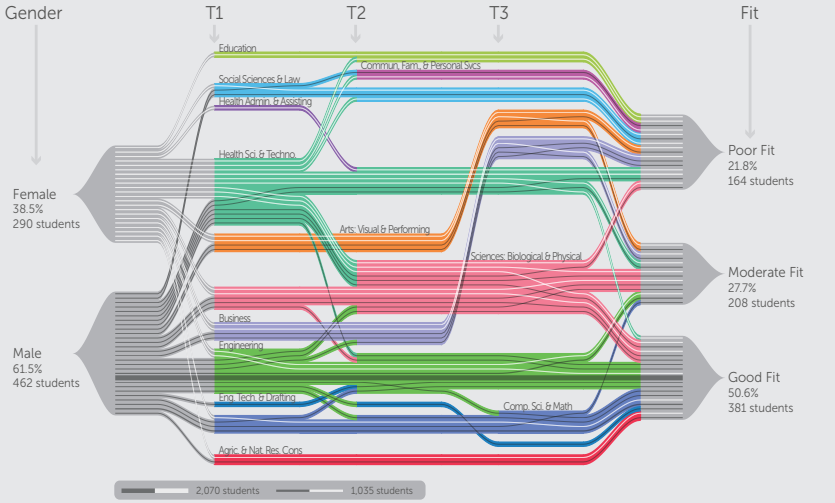
5,642 students took the top 25 most common paths.

It appears to be more common for women to go into liberal arts, health, and biological sciences, whereas high scoring men frequently choose engineering and other technical fields.



Students that were a good fit for computer science and engineering when they registered for the ACT's

752 students took the top 50 most common paths.

Most students who started in engineering and computer science, and didn't change to other majors, fell into the good fit score category indicated by ACT in their sophomore year. Students who started in other fields mostly received a poor or a moderate fit.

Let's do more to make engineering and computer science attractive choices for those individuals who are a good fit for those majors.