YSS/Significance Magazine Webinar

The Joy of Clustering

Sam Tyner, PhD

About Me

- Current: Postdoc with Center for Statistics and Applications in Forensic Evidence (CSAFE)
- Iowa State University Dept. of Statistics
  - PhD, December 2017
  - MS, May 2015
  - Advisor: Heike Hofmann
- YSS Runner-up in 2015 for “The Joy of Clustering”

The joy of clustering

Written by Samantha Tyner on 11 November 2015. Posted in Culture

In April 2014, FiveThirtyEight published a statistical analysis of the paintings of Bob Ross. Journalist Walt Hickey hand-coded hundreds of pieces of art in order to identify common themes. If that sounds like too much hard work, statistics PhD student Samantha Tyner describes an alternative machine-led approach that could make future art analysis a little bit easier.

If you watched public television in the US at any point in the 1980s or 1990s, then you have probably heard

Articles by date

▼ 2017 (10)
▼ March (3)
• E-sports, mind sports and the Olympics: What is a sport, anyway?
• A variation of the birthday problem - a reader responds
• What is the most popular birthday in England and Wales?
► February (3)
► January (4)
► 2016 (50)
► 2015 (75)
► 2014 (15)

Categories

Sports
Culture
Politics

https://www.statslife.org.uk/culture/2553-the-joy-of-clustering

Questions from Brian

- How did you come up with the idea for the article?
- Offer advice and guidance to prospective entrants
- How to craft an article that presents an accessible and engaging discussion and explanation of statistical ideas and methods?

Statistics 503 - Exploratory Methods and Visual Data Mining
Syllabus Spring 2015

Course Description: Approaches to finding the unexpected in data: data mining, pattern recognition and understanding. Emphasis is on data-centered, non-inferential statistics, for large or high-dimensional data, and topical problems. Simple graphical methods, as well as classical and computer-intensive methods applied in an exploratory manner.

- Started out as a class project!
- Professor encouraged me to publish somewhere

Di Cook, Professor
Econometrics & Business Statistics
Monash University, Melbourne, Victoria, AUS
dicook.github.io
@visnut

Project Idea

- Perusing FiveThirtyEight's data repository on Github
- https://github.com/fivethirtyeight/data
- "There's gotta be a faster way to do this…"

FiveThirtyEight

Politics Sports Science & Health Economics **Culture** Podcast: What

APR. 14, 2014 AT 10:21 AM

A Statistical Analysis of the Work of Bob Ross

By Walt Hickey

Filed under Happy Spreadsheets

Get the data on GitHub

Bob Ross in 1985. "THE JOY OF PAINTING," THE BOB ROSS NAME AND IMAGE ARE TRADEMARKS OF BOB ROSS INC. ALL RIGHTS RESERVED. USED WITH PERMISSION.

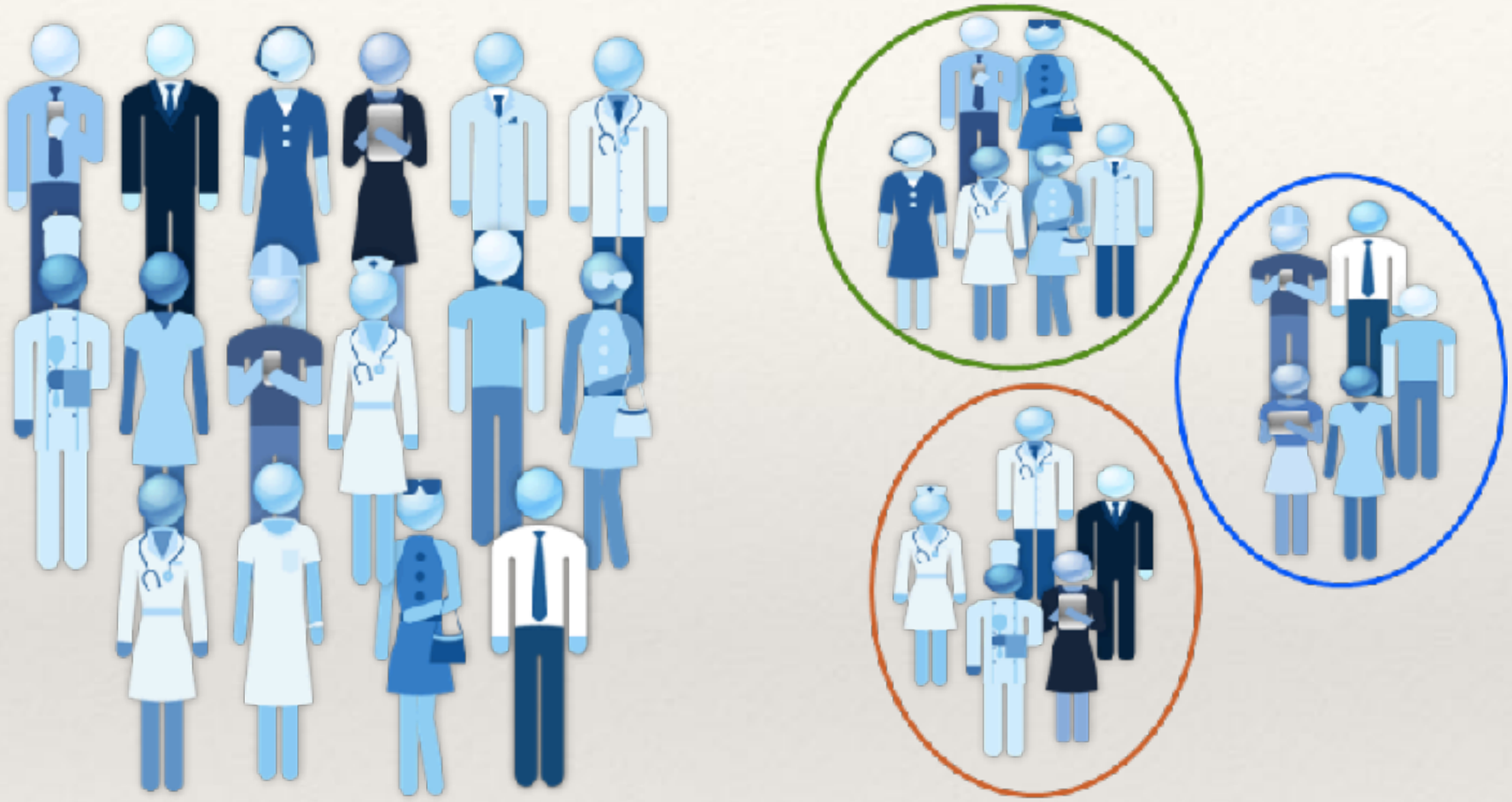
Bob Ross was a consummate teacher. He guided fans along as he painted "happy trees," "almighty mountains" and "fluffy clouds" over the course of his 11-year television career on his PBS show, "The Joy of Painting." In total, Ross painted 381 works on the show, relying on a distinct set of elements, scenes and themes, and thereby providing thousands of data points. I

https://fivethirtyeight.com/features/a-statistical-analysis-of-the-work-of-bob-ross/

Accessible & Engaging

- Many people with basic cable in the 80s & 90s in the US are familiar with the show "The Joy of Painting"
- (Also all episodes are now on Netflix in the US)
- People can imagine watching hundreds of episodes and hand-coding them.
- Can see advantage of a sped-up process
- Tell a story, show don't tell

Explaining Statistical Concepts

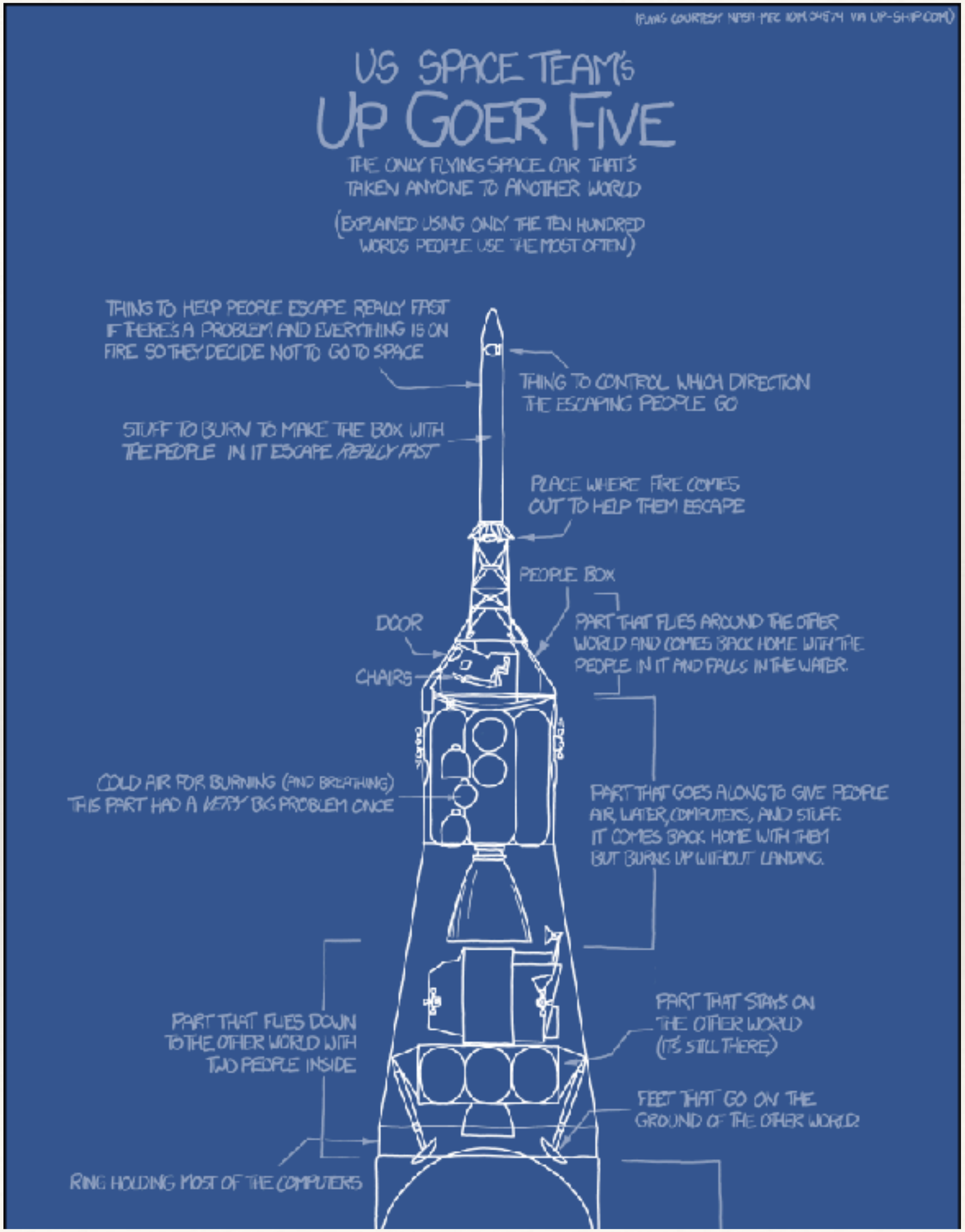


https://blogs.sas.com/content/subconsciousmusings/files/2016/05/clustering-based-on-similarities.png

- Clustering is fairly intuitive
- Find an analogy / example people are familiar with

https://xkcd.com/1133/

- Use appropriate language
- Avoid jargon
- “Ten hundred words people use the most often”



Conclusion

- What do YOU think is interesting?
- Tell a complete story and avoid jargon
- Have fun!