

Penn State Microbiome Center Presents

Title:
Data Visualizations and Immersive
Experiences:
Changing the Perspective of Your Data and
Providing New Insights

Name: Patrick M. Dudas
Affiliation: Institute for Computational and Data
Sciences

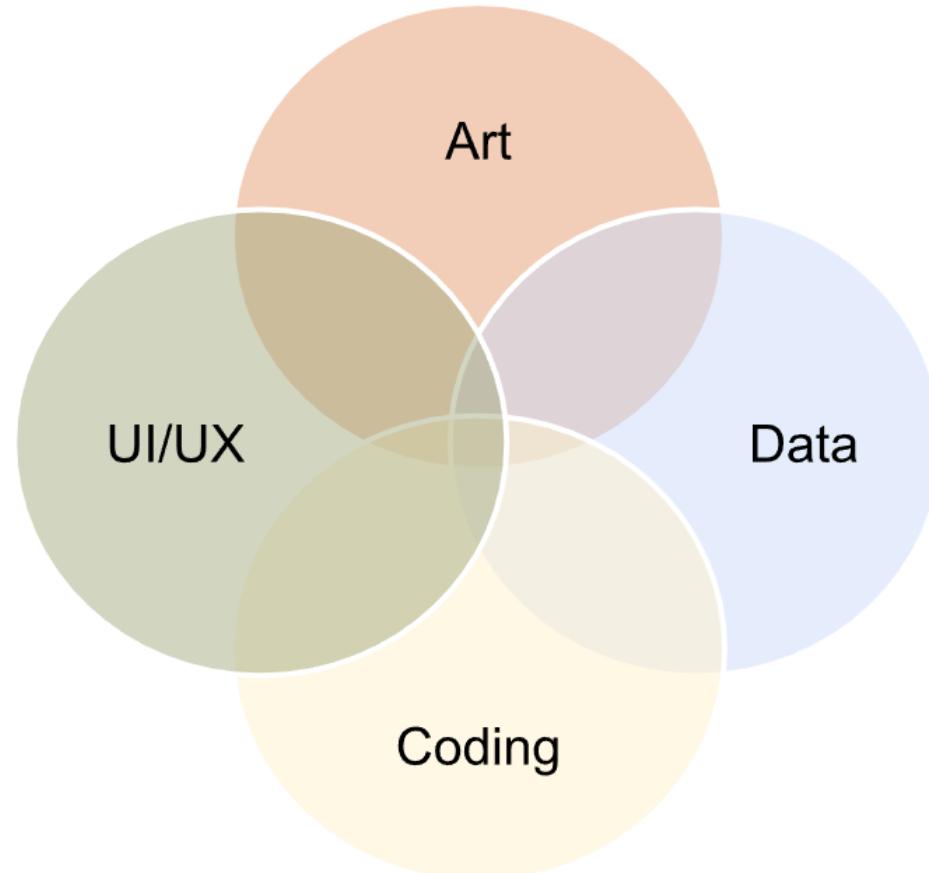
Who Am I?



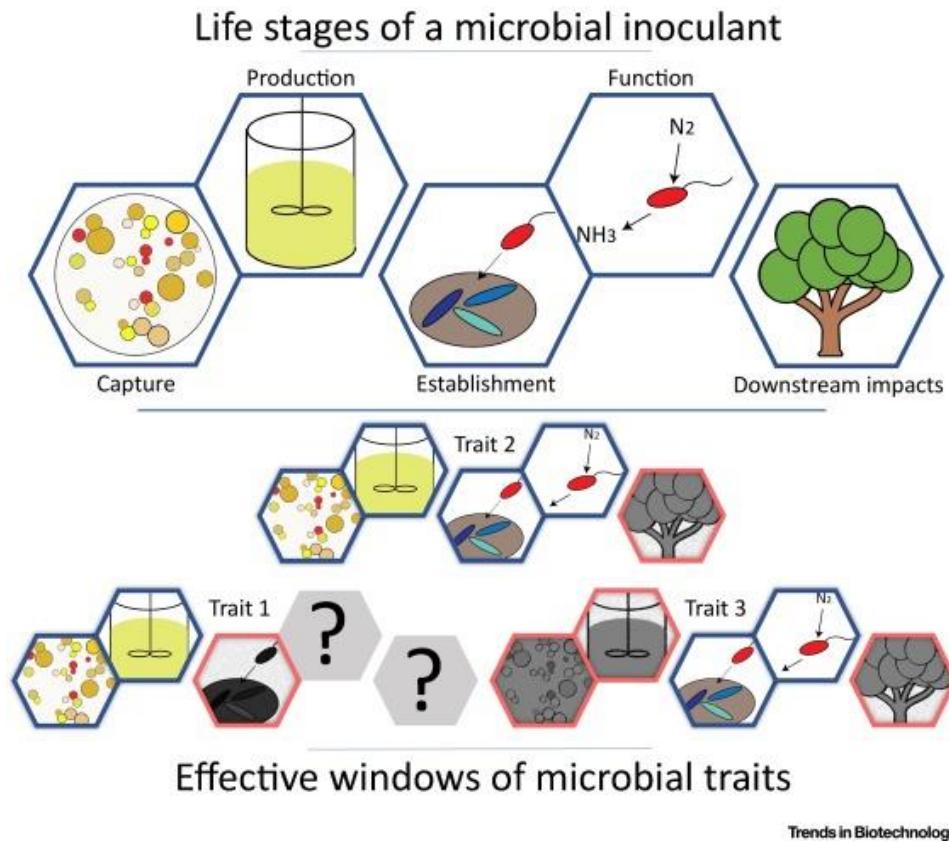
- Patrick M. Dudas, PhD
- pronouns = [he,him,his]
- Interim Director of the Center for Immersive Experiences
- Data Visualization Research & Development Engineer
 - Institute for Computational and Data Sciences
- Passions: Data and Storytelling Fluency



Data visualizations



Graphic Development



- Terrence Bell

- Kaminsky, L. M., Trexler, R. V., Malik, R. J., Hockett, K. L., & Bell, T. H. (2019). The inherent conflicts in developing soil microbial inoculants. *Trends in Biotechnology*, 37(2), 140-151.



Backstory

- A Jupyter Book Approach to Latent Dirichlet Allocation Understanding
- A 'Sourceful' Twist: Emoji Prediction Based on Sentiment, Hashtags and Application Source
- A Digital Fluency Framework to Support 21st-Century Skills
- Long-term effects of increased adoption of artemisinin combination therapies in Burkina Faso
- Bee survival: An applied network analytical strategy
- The ensemble mars atmosphere reanalysis system (emars) version 1.0
- Prevalence of Bovine Tuberculosis in India: A systematic review and meta-analysis

- Collaborations with researchers in Biology, Meteorology, Business, Infectious Diseases, Political Science, Veterinary and Biomedical Sciences, Entomology, Food Science, Chemistry, Physics, Engineering, Information Science, and Clinical Pharmacology



The Art of Cell Biology

October 15th, 2018

Hintz Alumni Center, Robb Hall

Emily Bell



Journals are Boring*

We Can Do Better

* Your research is not boring. Quite the contrary, your research is the most interesting research out there! I am referring to journals in general. They are boring. Definitely not your research.



Philosophical Transactions (1665-1678) - Royal Society



(215) Num. 13.
PHILOSOPHICAL
TRANSACTIONS.

Munday, June 4. 1666.

The Contents.

Certain Problems touching some Points of Navigation. Of a new Contrivance of Wheel-Barometer, much easier to be prepar'd, than others. An account of Four Sun's which lately appear'd in France; and of two, unusually posited, Rainbows, seen in the same Kingdom. A Relation of an Accident, by Thunder and Lightning, in Oxford. An Experiment, to examine, what Figure or Celerity of Motion begetteth or increaseth Light and Flame. Some Considerations touching a Letter in the Journal des Scavans of May 24. 1666.

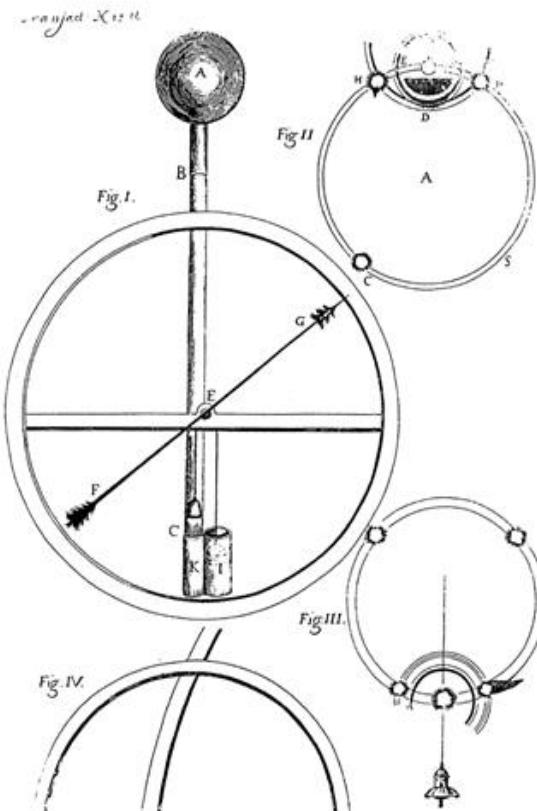
Certain Problems
Touching some Points of Navigation.



These Problems are presented by the Learned and Industrious Nicolaus Mercator, for the advancing of that Excellent and Beneficial Science, Navigation, as follows:

The line of Artificial Tangents, or the Logarithmical Tangent-line, beginning at 45 deg, and taking every half degree for a whole one, is found to agree pretty near with the Meridian-line of the Sea-Chart; they both growing, as it were, after the same Proportion. But the Table of Meridional degrees being calculated only to every Sexagesimal minute of a degree, shews some small difference from the said Logarithmical Tangent-line. Hence it may be doubted, whether that difference do not arise from that little error, which is committed by calculating the Table of Meridional degrees only to every minute.

Mr. G g



352 Years Later...

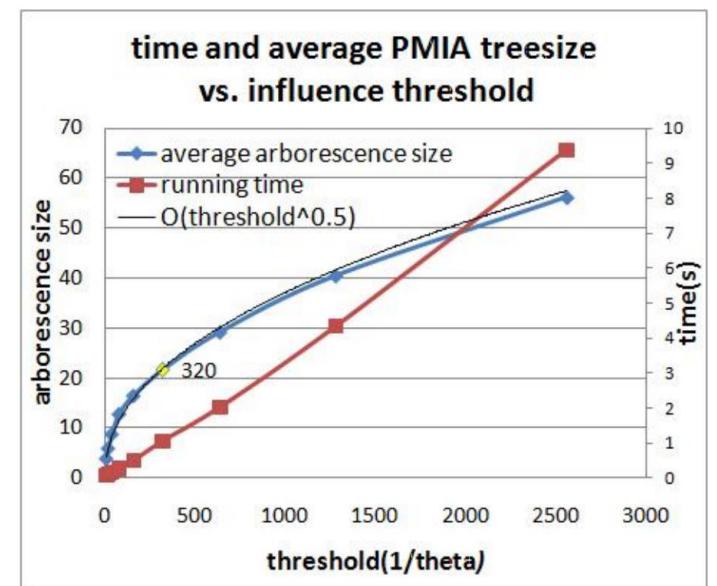
Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks

ABSTRACT

Influence maximization, defined by Kempe, Kleinberg, and Tardos (2003), is the problem of finding a small set of seed nodes in a social network that maximizes the spread of influence under certain influence cascade models. The scalability of influence maximization is a key factor for enabling prevalent viral marketing in large-scale online social networks. Prior solutions, such as the greedy algorithm of Kempe et al. (2003) and its improvements are slow and not scalable, while other heuristic algorithms do not provide consistently good performance on influence spreads. In this paper, we design a new heuristic algorithm that is easily scalable to millions of nodes and edges in our experiments. Our algorithm has a simple tunable parameter for users to control the balance between the

close social circle of families, friends, and co-workers. Research shows that people trust the information obtained from their close social circle far more than the information obtained from general advertisement channels such as TV, newspaper and online advertisements [16]. Thus many people believe that word-of-mouth marketing is the most effective marketing strategy (e.g. [15]).

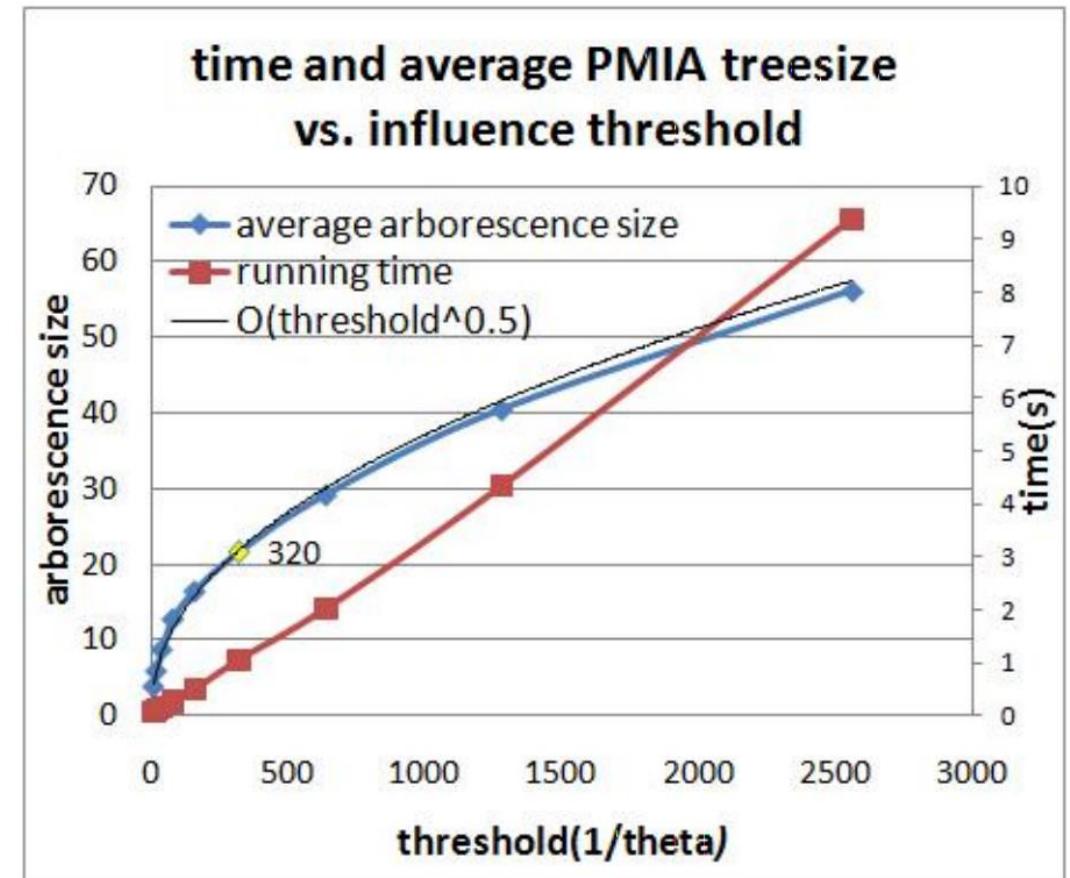
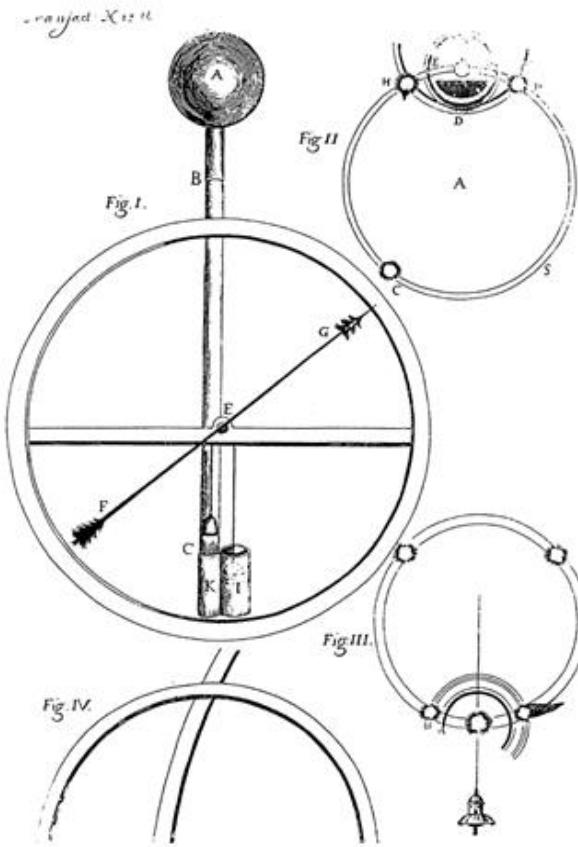
The increasing popularity of many online social network sites, such as Facebook, Myspace, and Twitter, presents new opportunities for enabling large-scale and prevalent viral marketing online. Consider the following hypothetical scenario as a motivating example. A small company develops a cool online application and wants to market it through an online social network. It has a limited budget such that it can only select a small number of initial



Chen, W., Wang, C., & Wang, Y. (2010, July). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1029-1038).

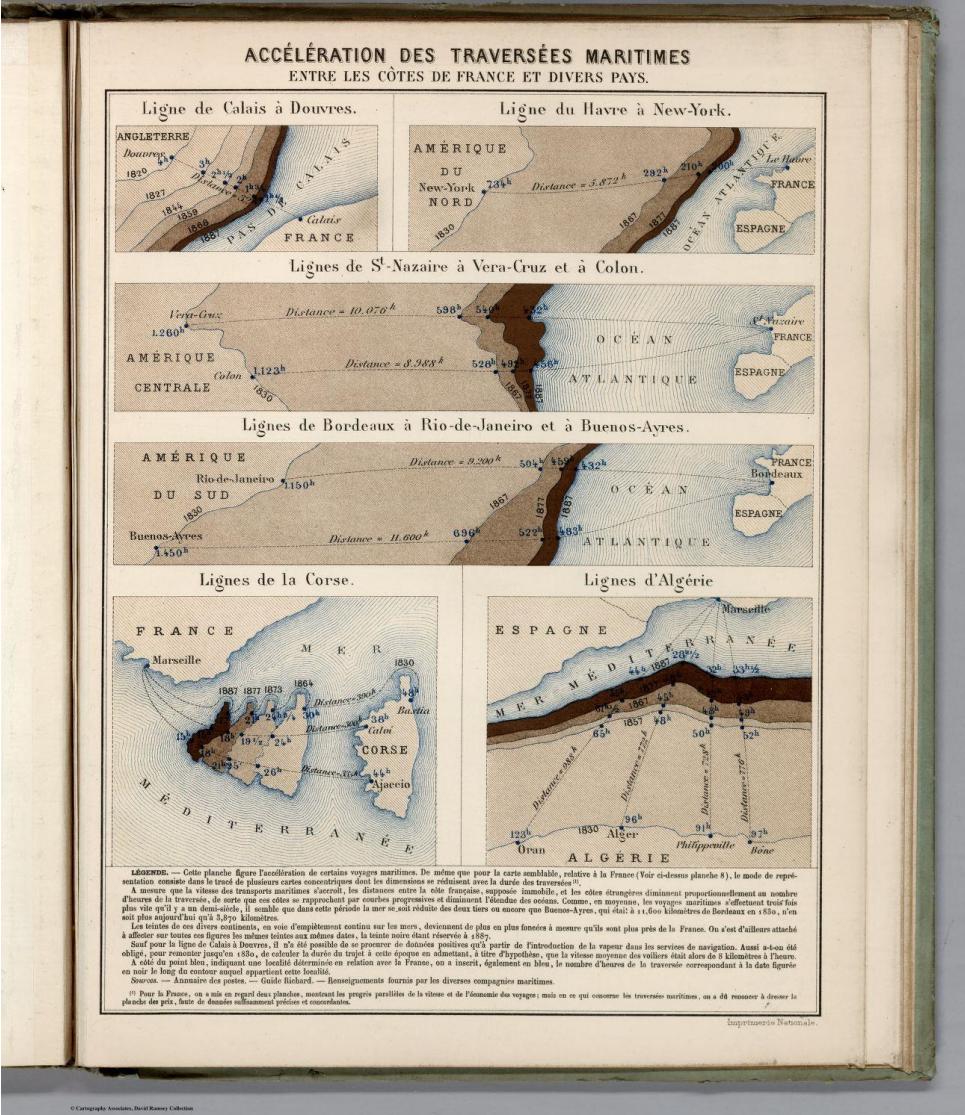


Can we not do better?



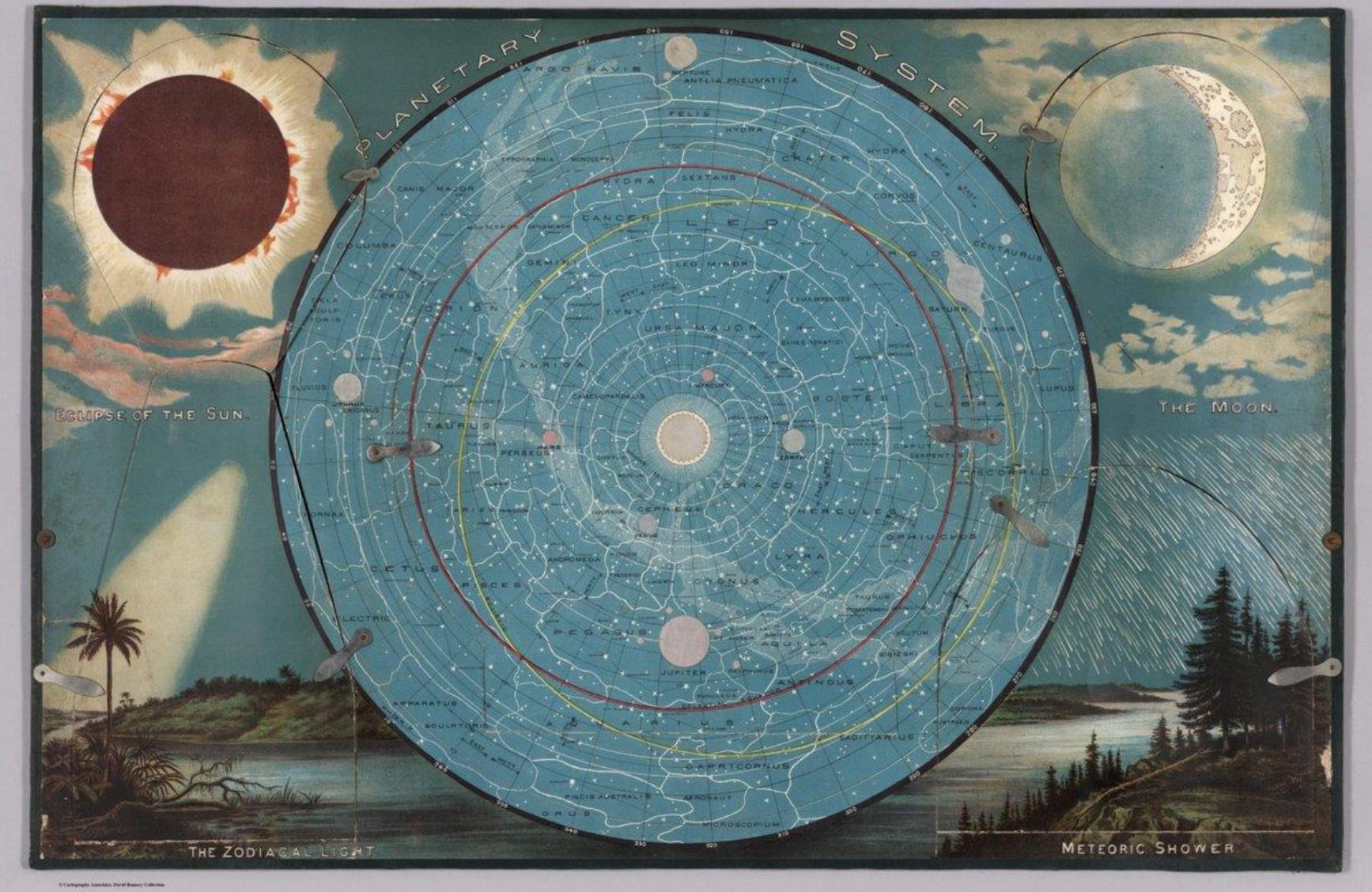
**The following pieces are
from the
David Rumsey Map Center |
Stanford Libraries**





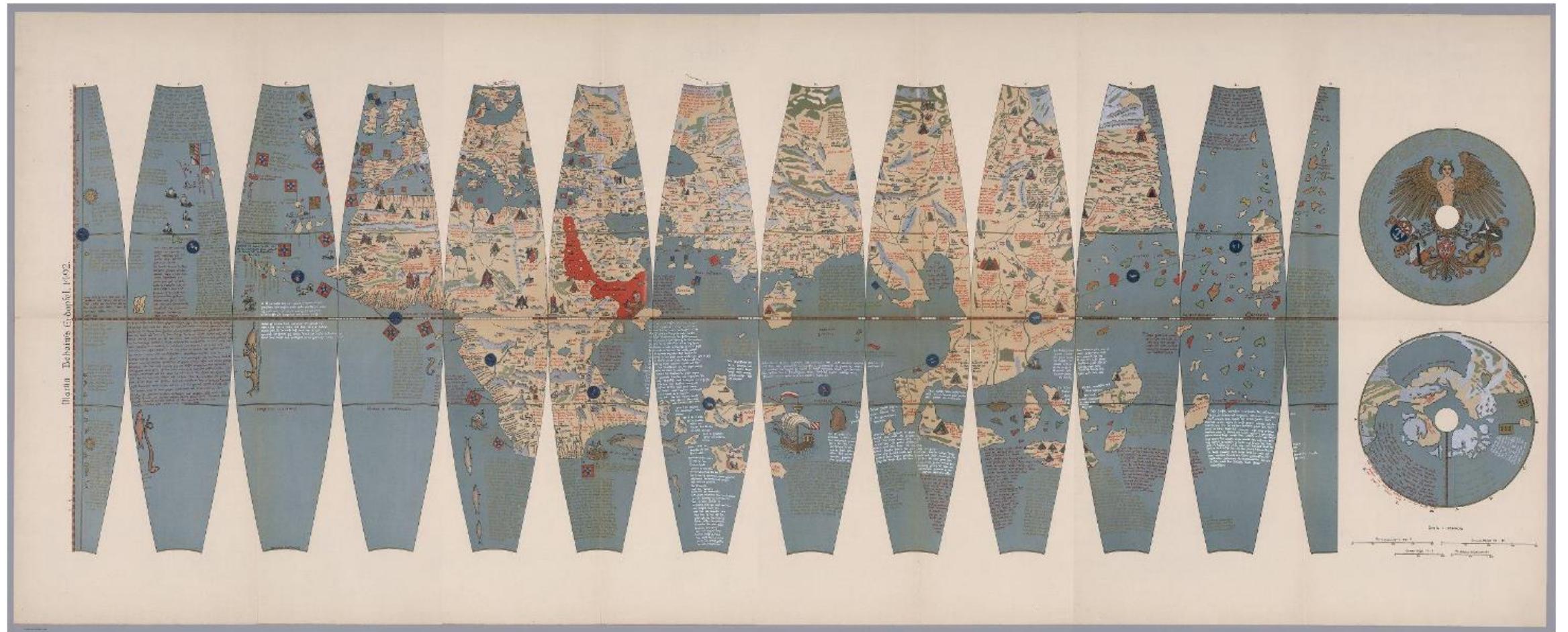
Cheysson, Émile 1888 "Acceleration of the maritime crossings between the French coasts and various countries."





Levi Walter Yaggy's 1887 "Geographical Study"





Behaim, Martin – 1492 “Martin Behaim's Erdapfel, 1492”



PennState

Institute for Computational
and Data Sciences

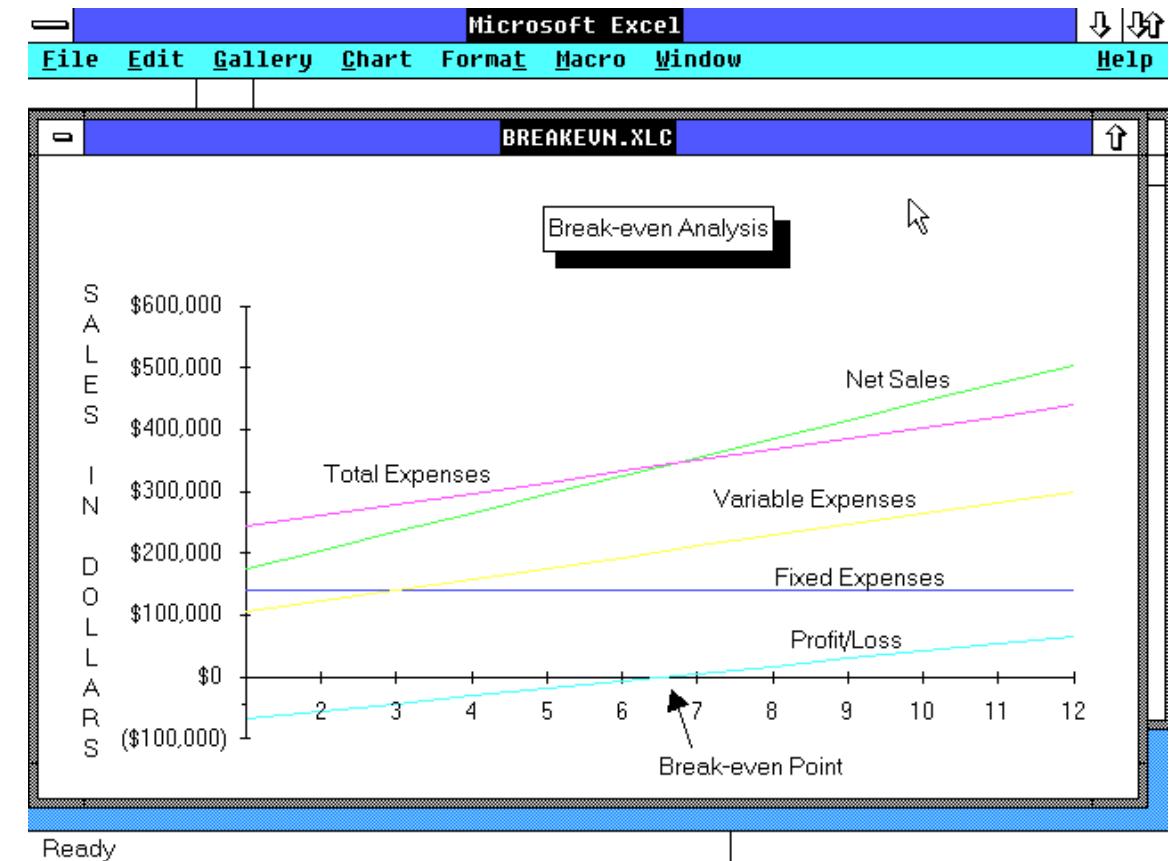
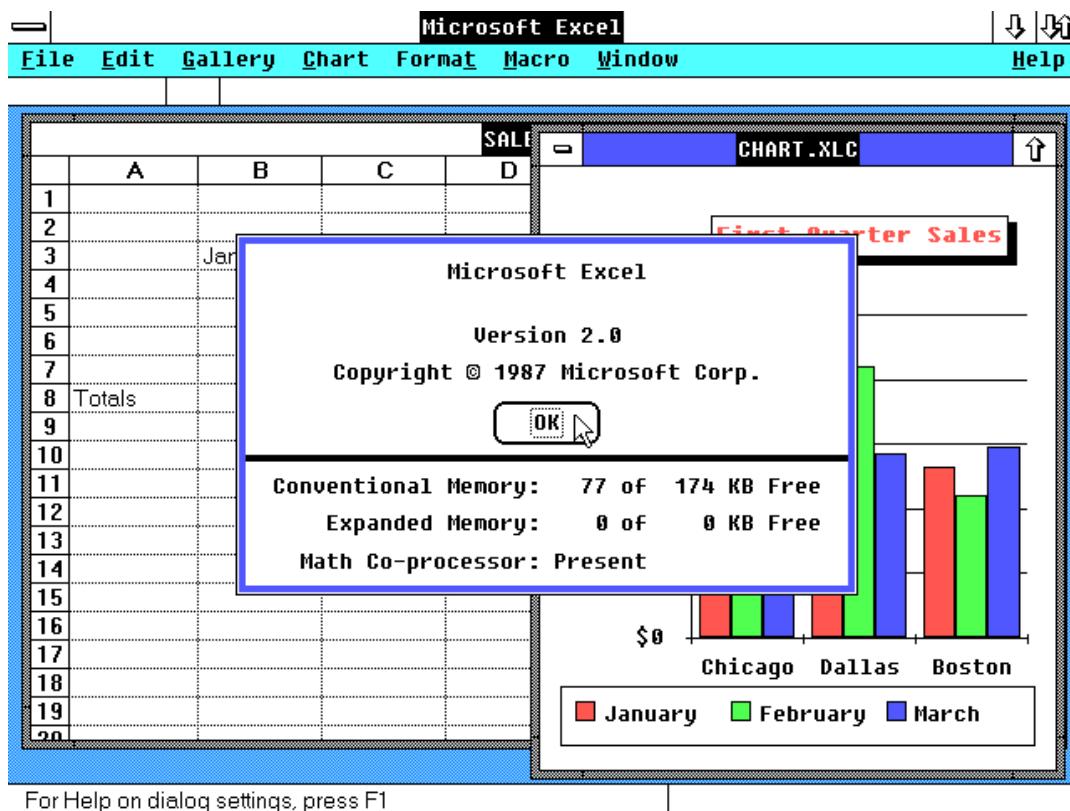
pmd19

Slide:14



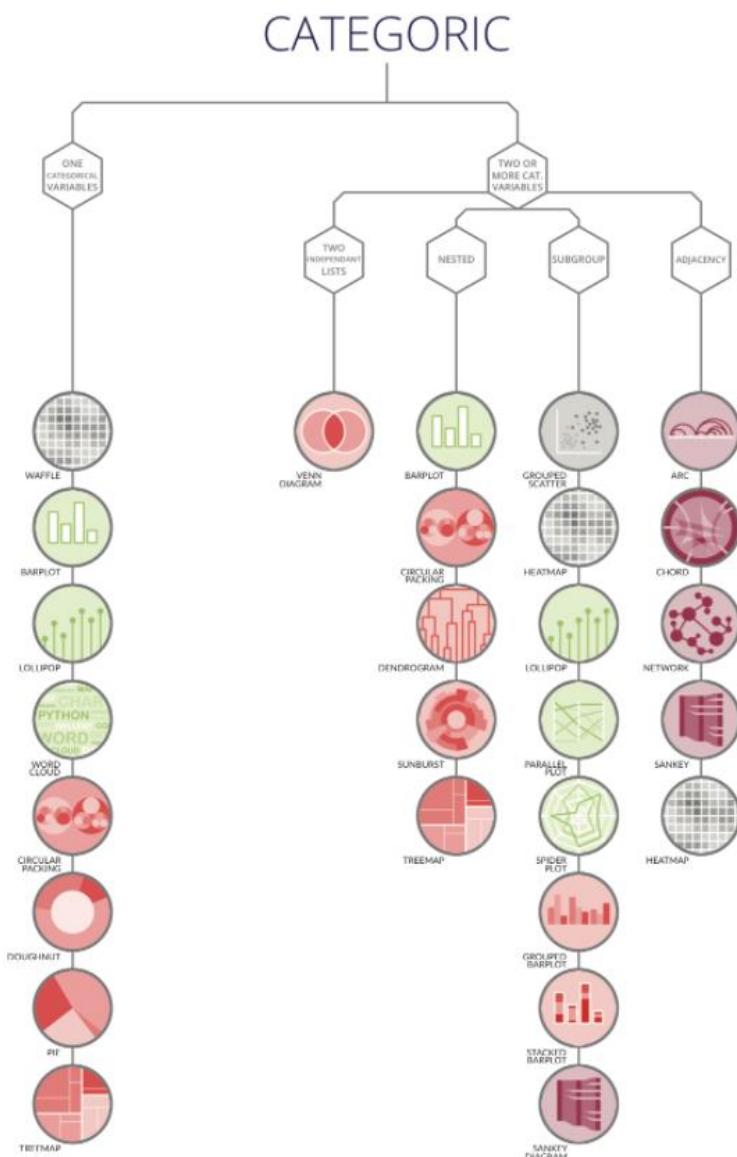
So, what happened?

Then Excel happened....



That being said..

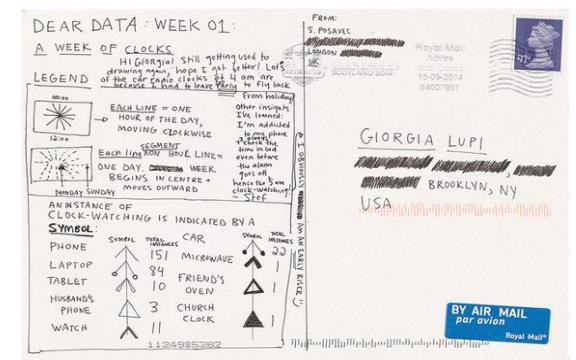
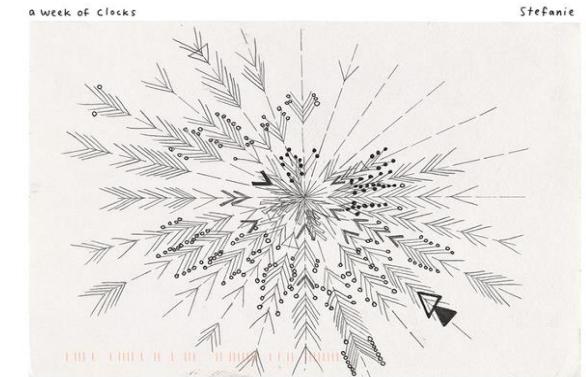
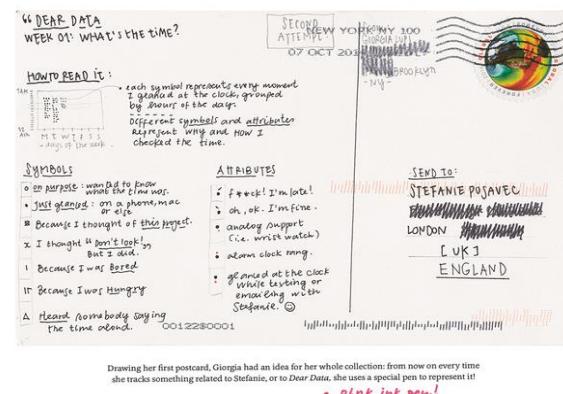
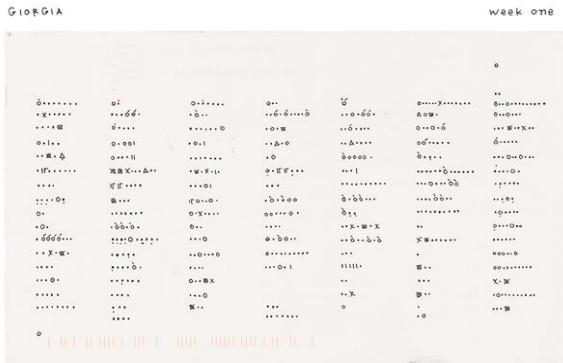
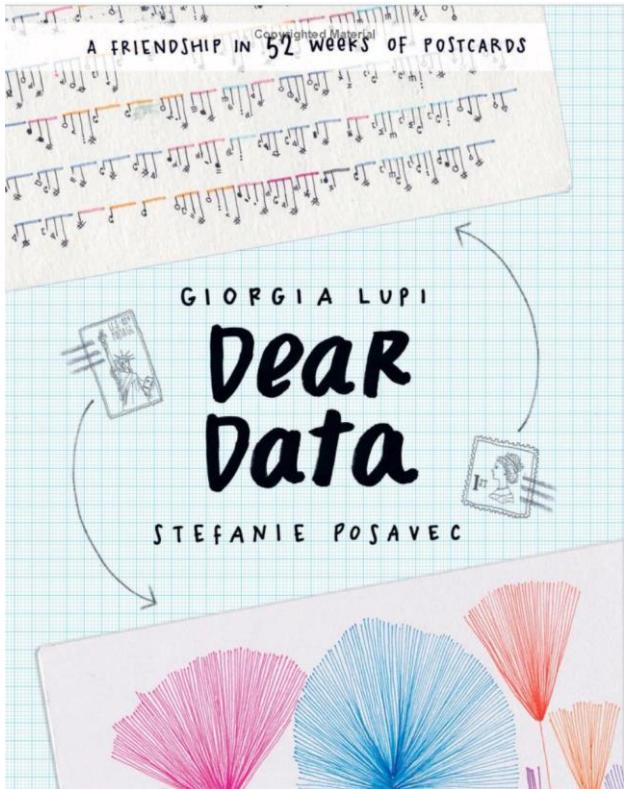




From Data to Viz, by Yan Holtz and Conor Healy, <https://www.data-to-viz.com/>



Dear Data – Giorgia Lupi & Stefanie Posavec



Lupi, G., & Posavec, S. (2016). *Dear data*. Chronicle books.





Kepler.gl by Shan He

<https://www.informationisbeautifulawards.com/showcase/3082-kepler-gl>



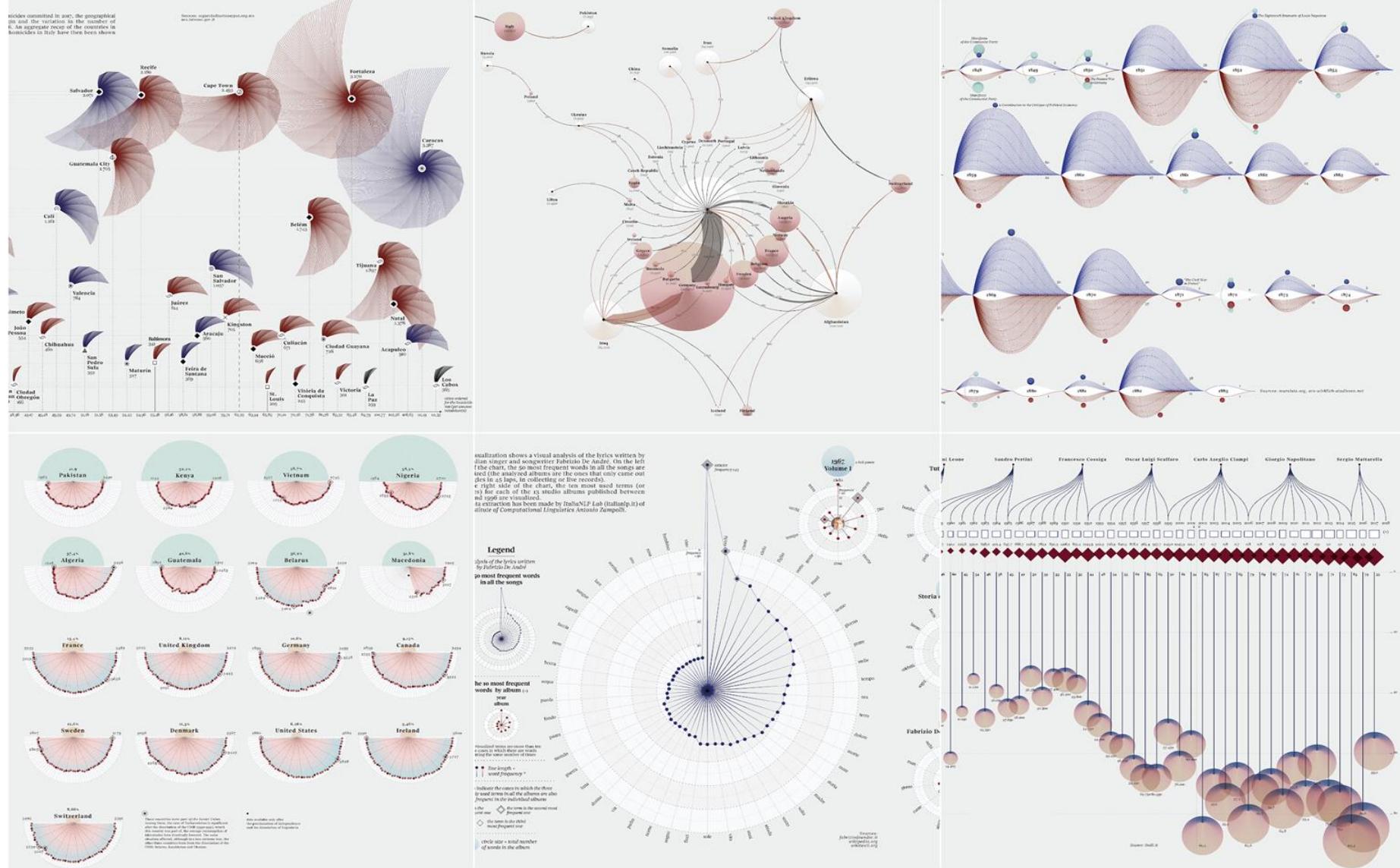
PennState

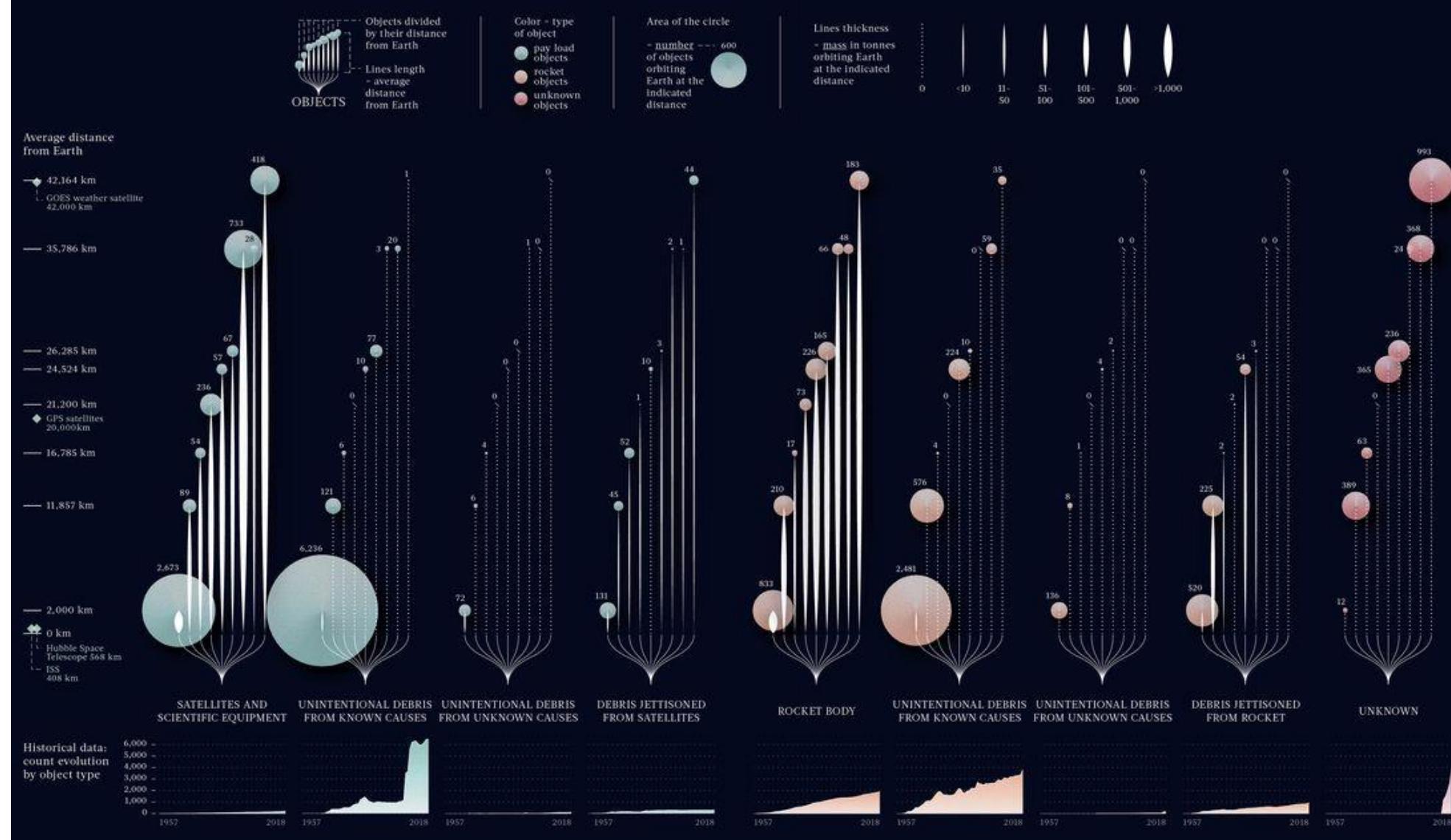
Institute for Computational
and Data Sciences

pmd19

Slide:20







Space Junk by Credits Art Editor: Joe Eden; News Editor: Jason Goodyer; Data visualization by Federica Fragapane
<https://www.informationisbeautifulawards.com/showcase/4268-space-junk>



**Interesting... but what
exactly are data
visualizations**



Reader- vs. Author-Driven

Reader-Driven

No prescribed ordering

No messaging

Free interactivity

Author-Driven

Linear ordering of scenes

Heavy messaging

No interactivity

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6), 1139-1148.



Computational Essays

- 1. Ordinary text.
 - a. context and motivation
 - 2. Computer input.
 - 3. Computer output.
-
- “And the crucial point is that these all work together to express what’s being communicated.”

<http://blog.stephenwolfram.com/2017/11/what-is-a-computational-essay/>

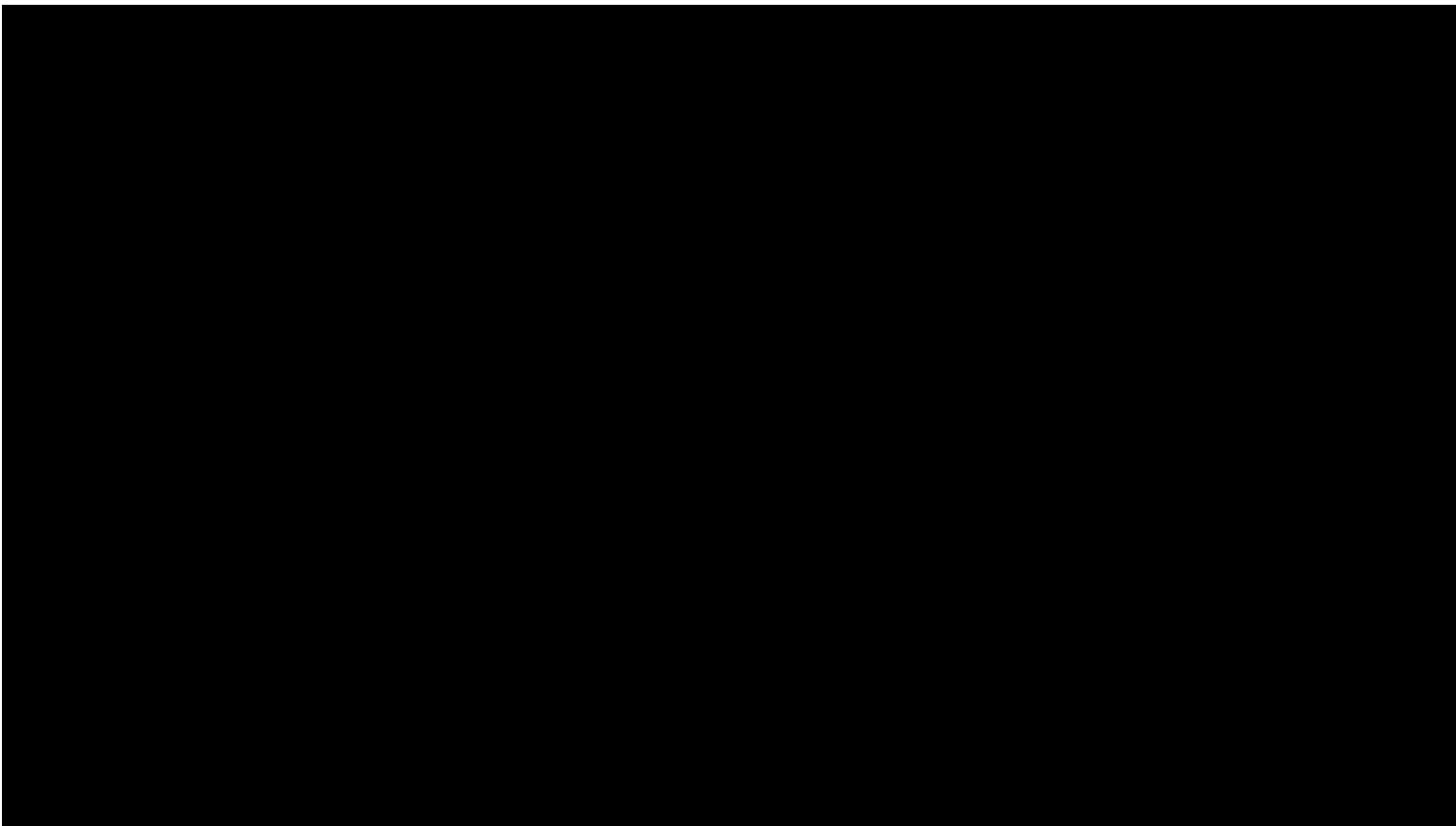


Interactive Notebook Examples

- Jupyter - <http://jupyter.org/>
- Google Collaborator - <https://colab.research.google.com>
- Observables - <https://observablehq.com/>



How I Envision Future Publications



Venkit, Pranav Narayanan, and Patrick Dudas. "A Jupyter Book Approach to Latent Dirichlet Allocation Understanding." *4th Workshop on Visualization for AI Explainability*, IEEE VIS, 18 Oct. 2021.



**This is great, but what
are my options?**



high-level introductions to various software packages



various graphical packages/libraries

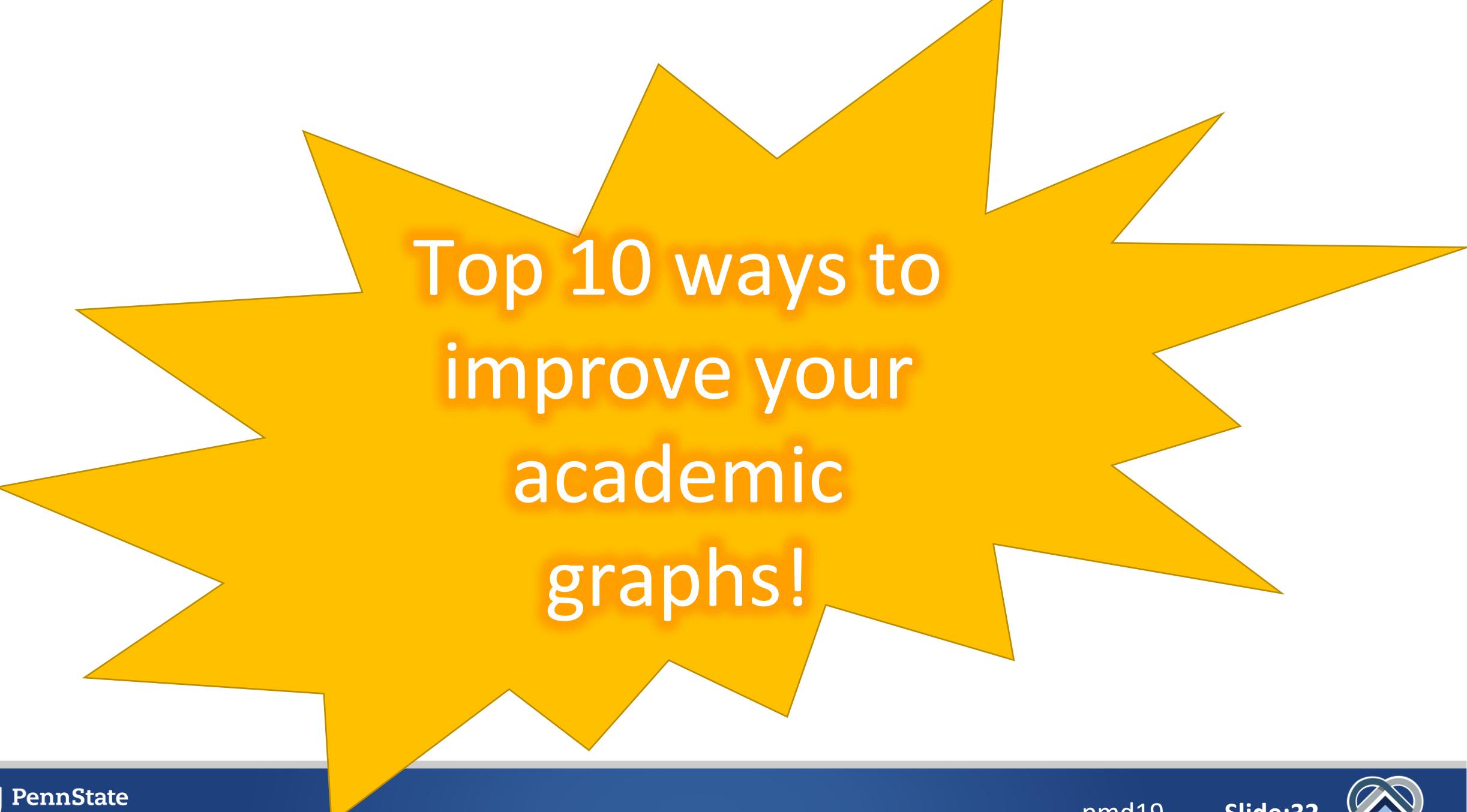
- Tableau
 - Standalone application (\$)
- Power BI
 - Included with Microsoft License
- ggplot2
 - R library
- Plot.ly
 - Python library
- vega-lite
 - JavaScript
- Matplotlib
 - Python
- D3.js
 - personal favorite (JavaScript)
 - biggest learning curve



Neat, can you just provide a few tips?

From: Taste of Methodology Workshop, Dec 13, 2018, Population Research Institute





Top 10 ways to
improve your
academic
graphs!



#10 – understand file formats

JPEG vs. PNG

- JPEG is lossy
- PNG is lossless

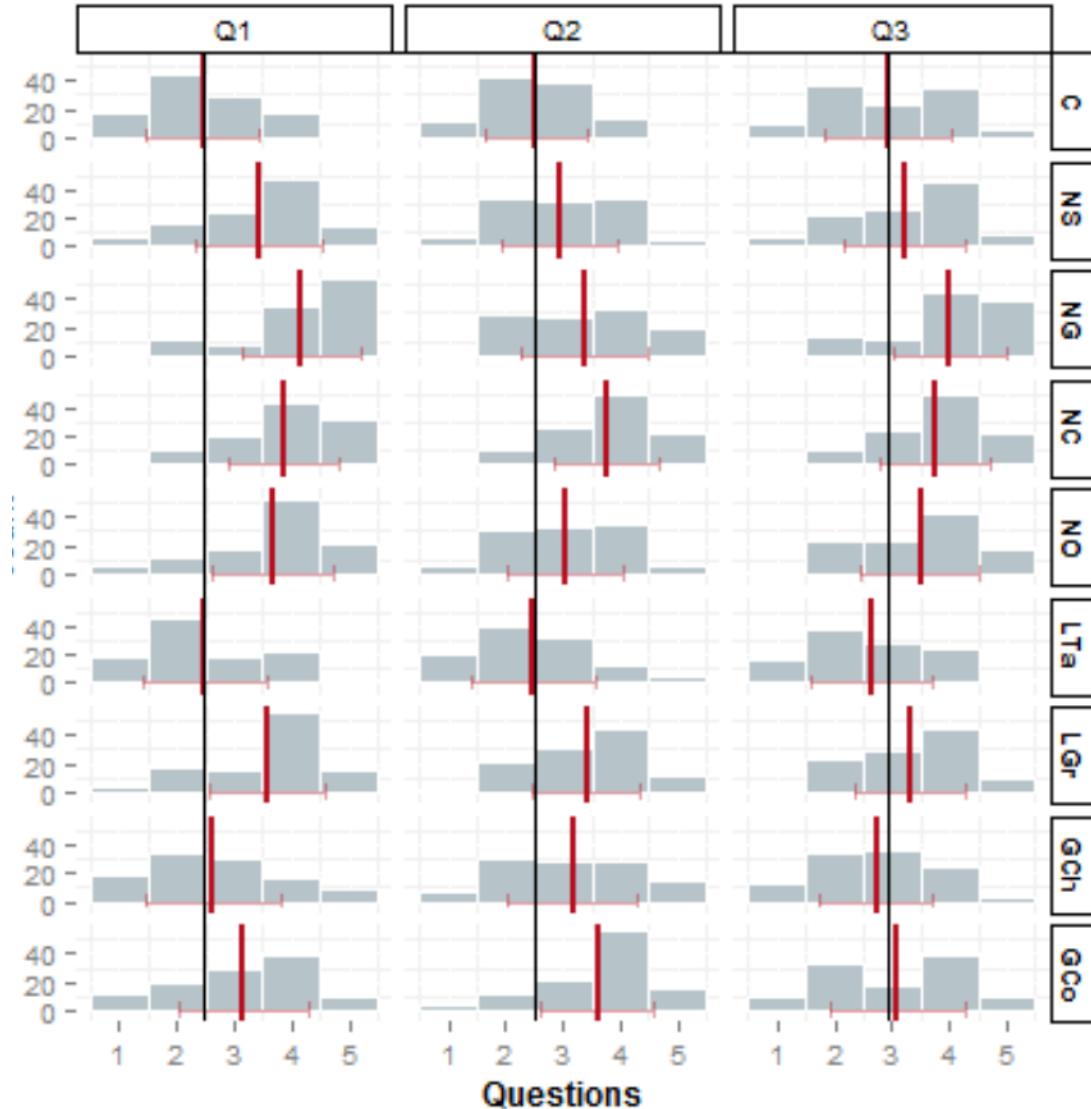
PNG vs. SVG (EPS/PDF)

- PNG is raster
- SVG is vector

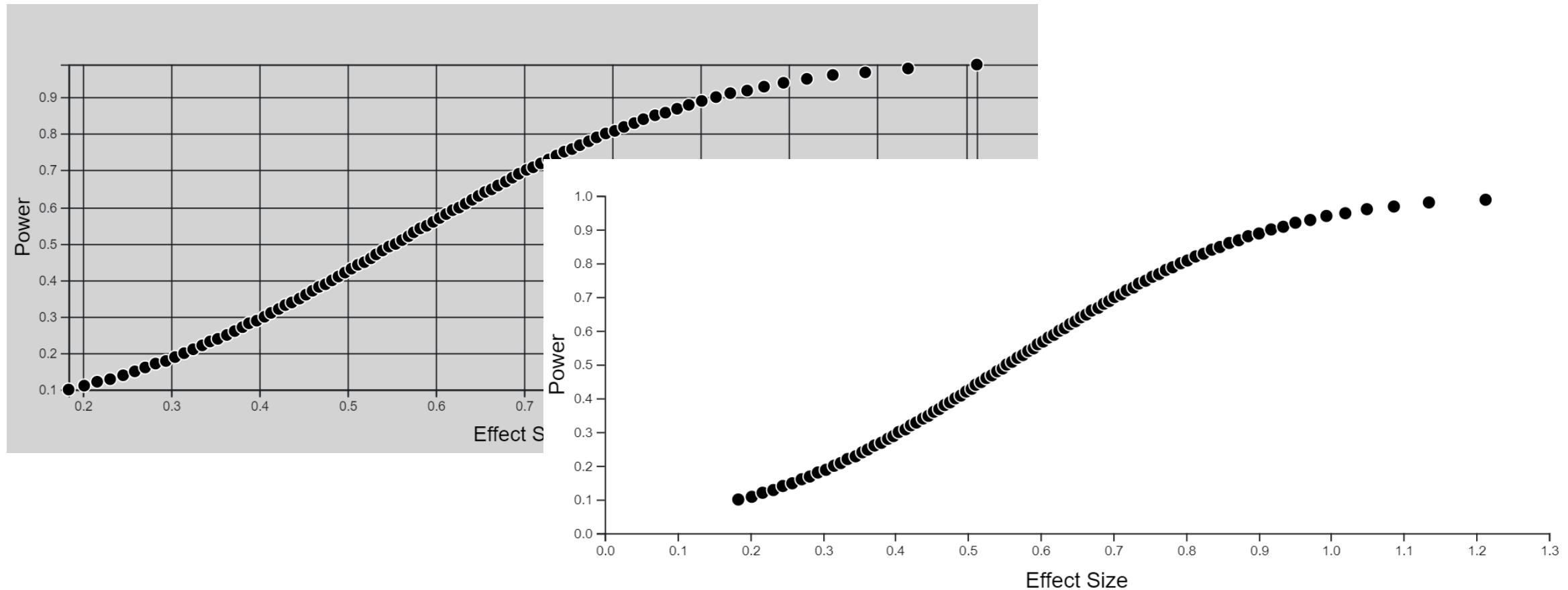


#9 – play “the game”

- This is a game I entirely made up.
- First, create your graph and caption for your publication.
- Then give only the graph to someone outside your field of research.
 - Have this person write a sentence or two about what they see.
- If your caption and their caption are completely different
 - Redo the figure.



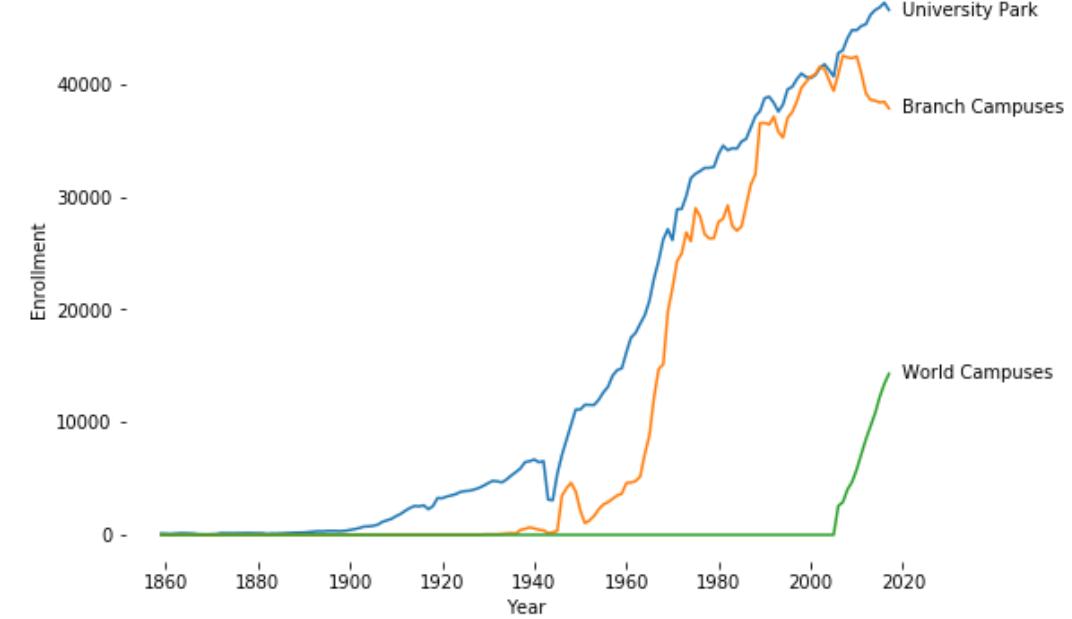
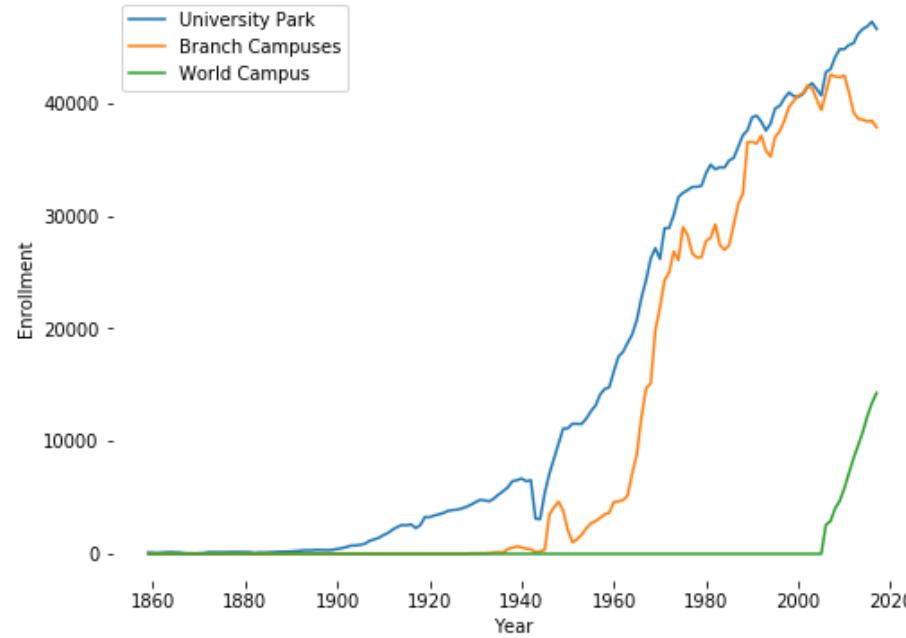
#8 - avoid “Chartjunk”



Tufte, E. R. (2006). Beautiful evidence (Vol. 1). Cheshire, CT: Graphics Press.



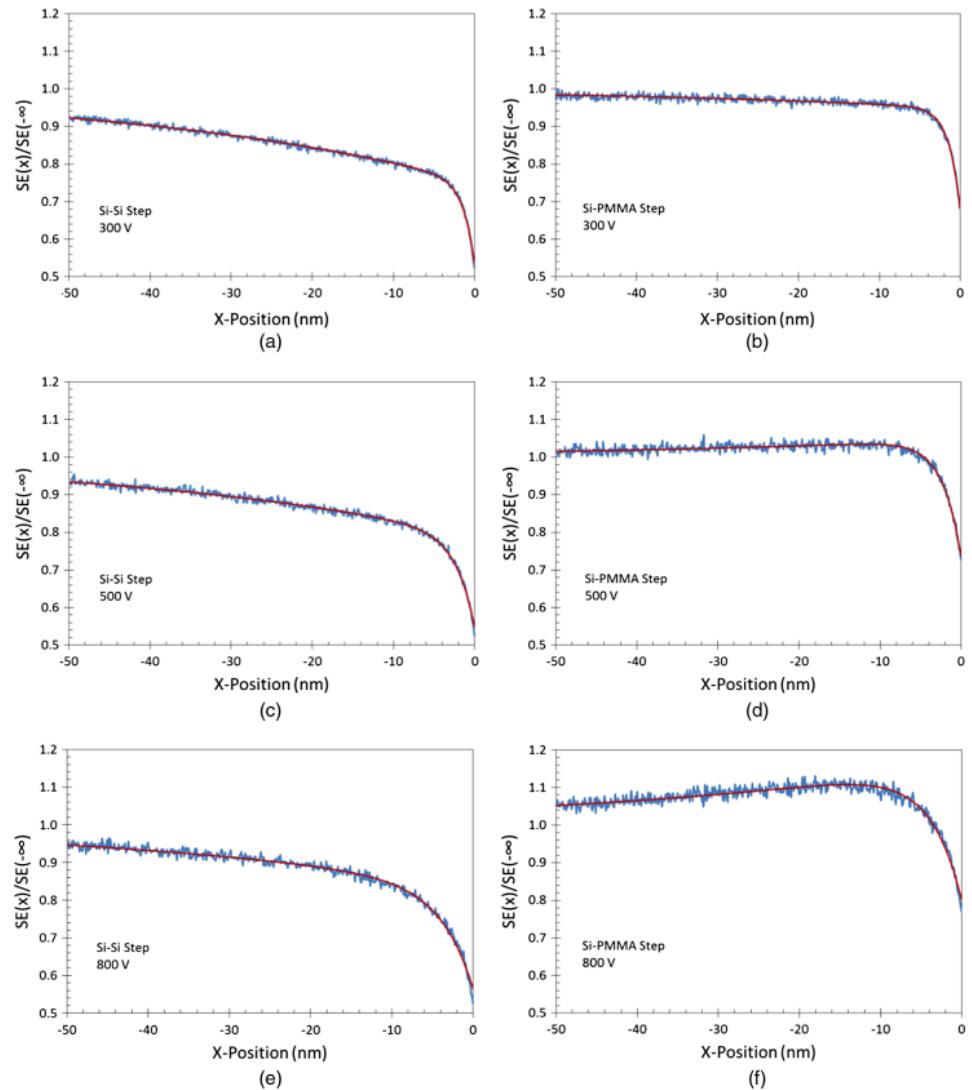
#7 - add labels to data, not just to the graph



Mack, C. A. (2014). How to write a good scientific paper: figures, part 2. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, 13(1), 010102.



#6 – when comparing plots, use the same x/y axis



Mack, C. A. (2014). How to write a good scientific paper: figures, part 2. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, 13(1), 010102.



#5 – axis/tick labels should always be the same font size as the paper font size (if not larger)

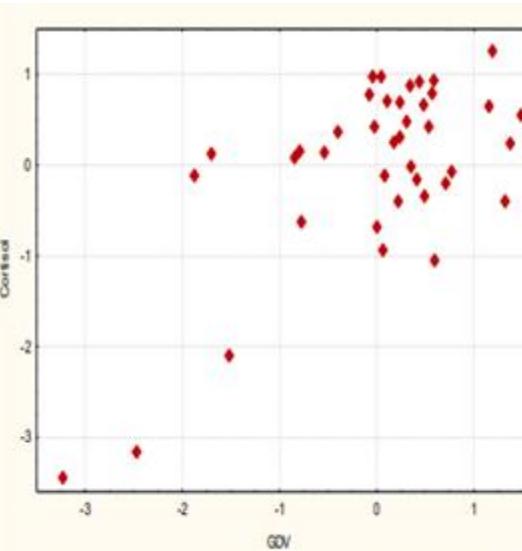
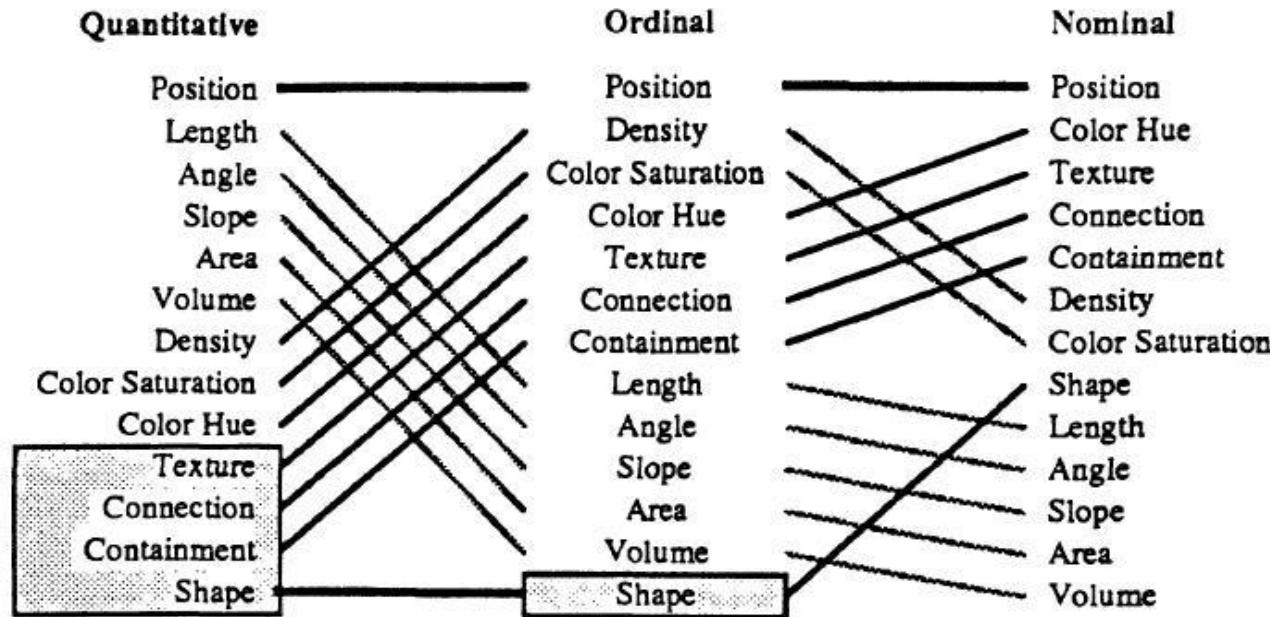


Fig. 4. Canonical correlation between GDV parameters (axis X) and serum Cortisol (axis Y)

Babelyuk, V. Y., Gozhenko, A. I., Dubkova, G. I., Korolyshyn, T. A., Kikhtan, V. V., Babelyuk, N. V., ... & Popovych, I. L. (2017). Causal relationships between the parameters of gas discharge visualization and principal neuroendocrine factors of adaptation. *Journal of Physical Education and Sport*, 17(2), 624.



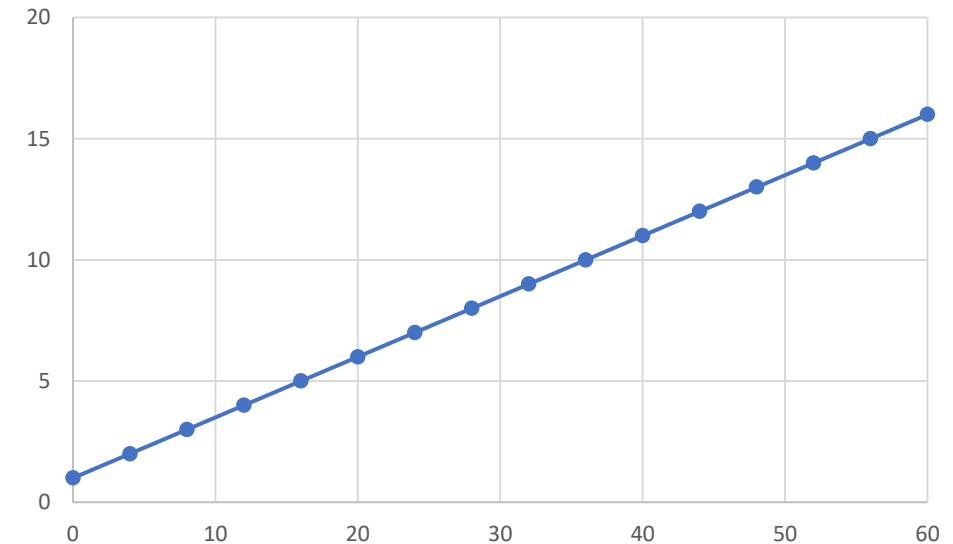
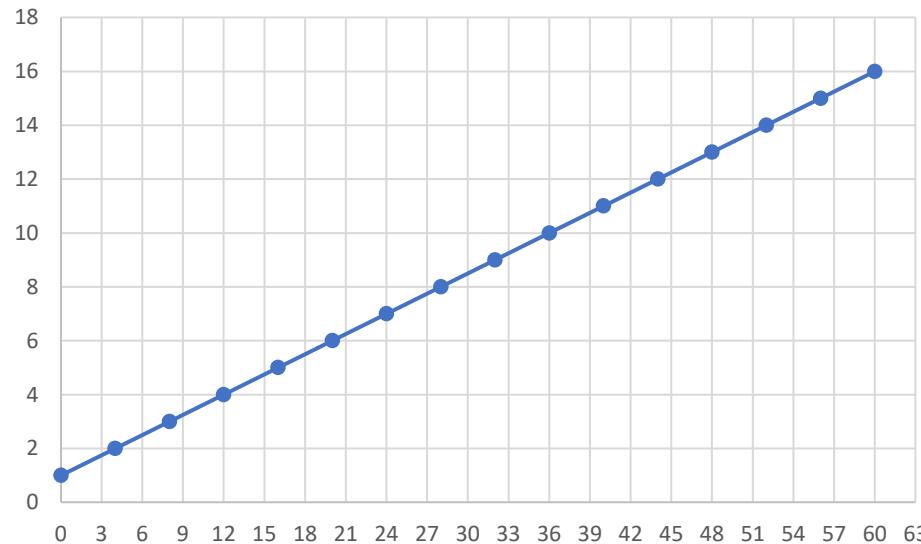
#4 – How should I encode my data?



Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387), 531-554.



#3 – Use sensible axis tick labels



#2 – picking the right spatial display for the data

- The Data Visualisation Catalogue
 - <https://datavizcatalogue.com/>
- R Graph Gallery
 - <https://www.r-graph-gallery.com/>
- From Data to Viz
 - <https://www.data-to-viz.com/>
- Xenographics
 - <https://xeno.graphics/>



#1 – understand that “color is important”

- Do not use the default colormap (rainbow)
- Sequential: One color, quantitative data varying from low to high.

- Diverging: Two colors, used to highlight deviation from a median.

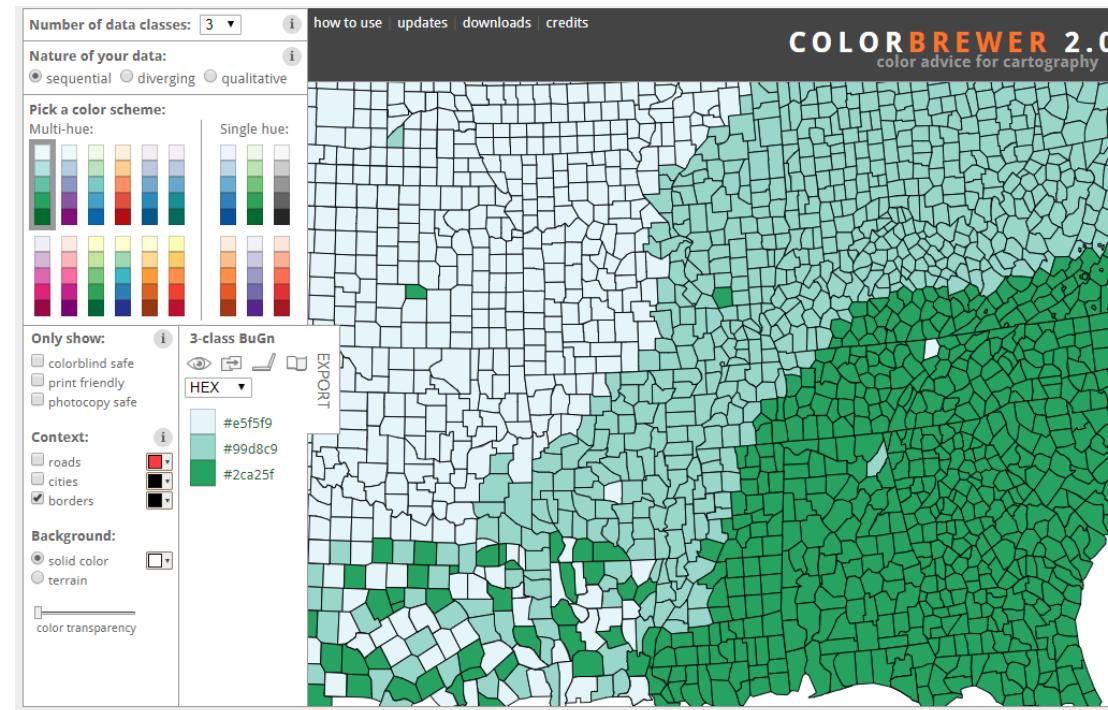
- Qualitative: rapid variation of colors for discrete or categorical data.
- Use HSL vs. RGB


Rougier, Nicolas P., Michael Droettboom, and Philip E. Bourne. "Ten simple rules for better figures." *PLoS computational biology* 10.9 (2014): e1003833.



one big takeaway

- <http://colorbrewer2.org/>



**OK, so how can you
help?**

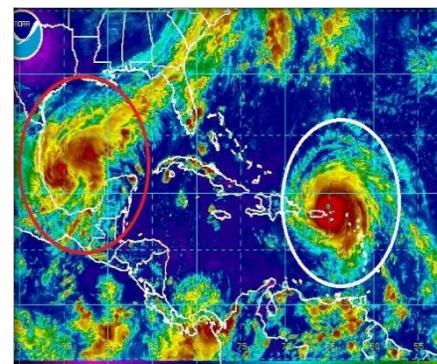
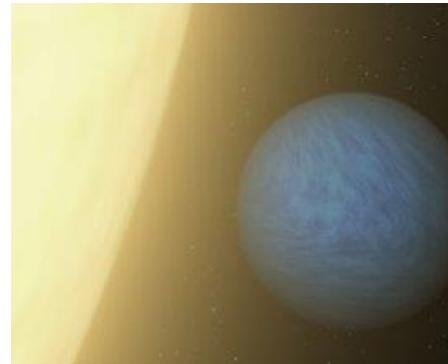


Institute for Computational and Data Sciences

Scholarly community



Interdisciplinary research



World-class cyberinfrastructure



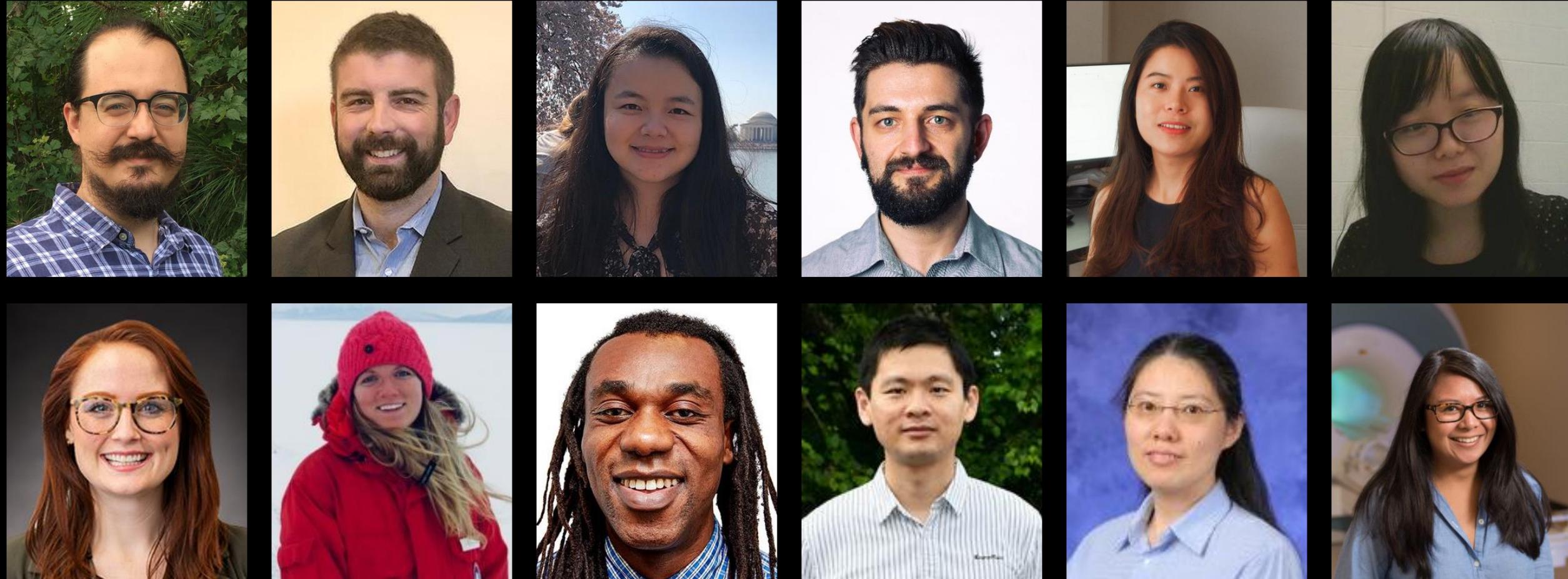


**34 ICDS Co-Hires in 21 Departments
[8 Colleges]**

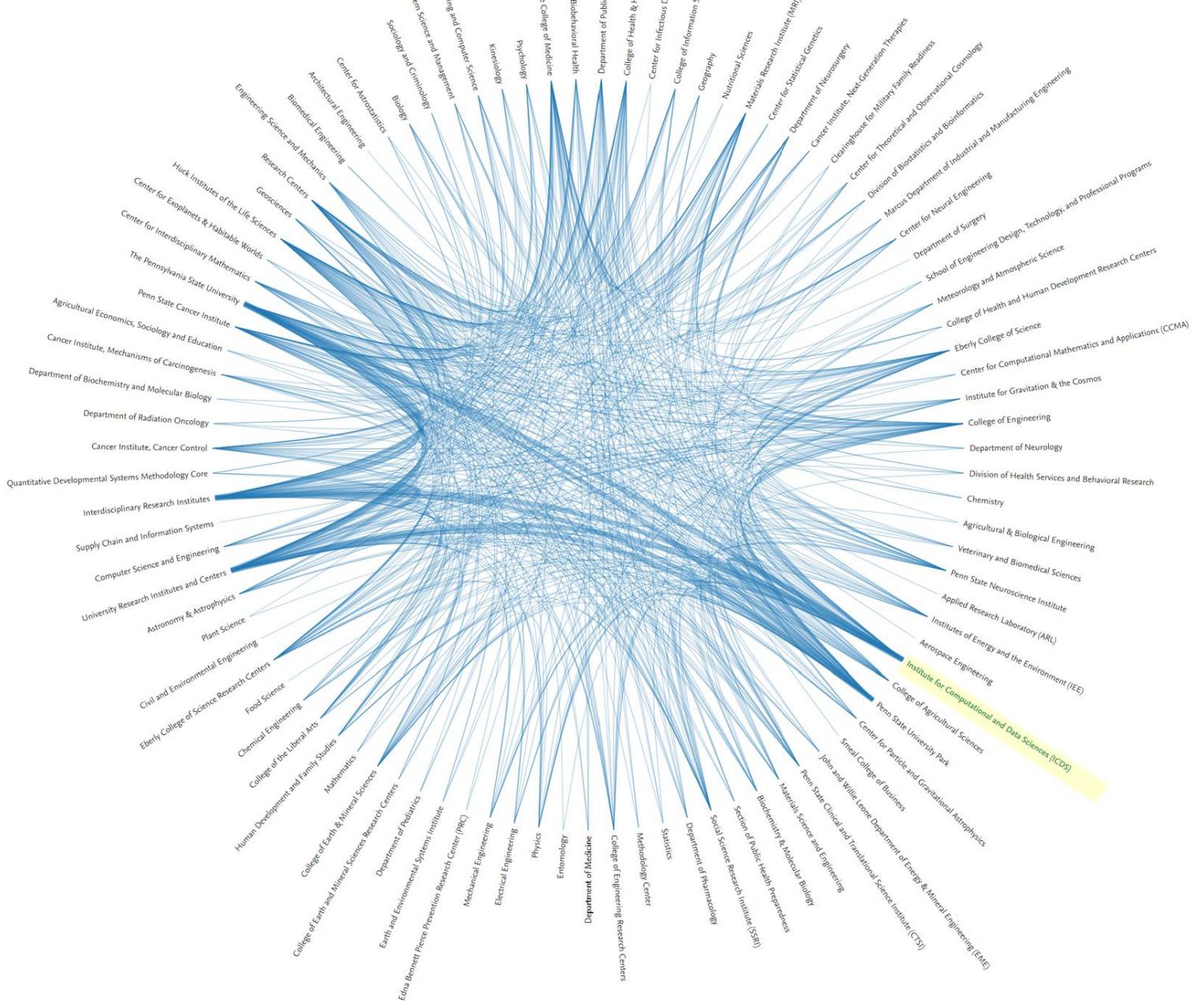


**New Co-Hires
in 2020**

2020 ICDS Associates & Affiliates



Interdisciplinarity

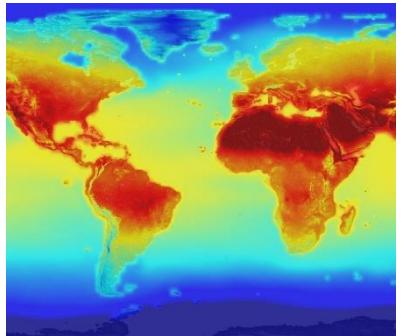


ICDS Collaborations (Units with 15 or more shared publications) with Penn State Internal units
2020-present

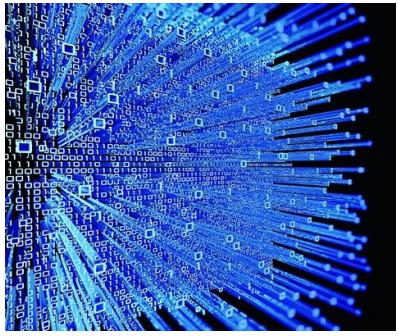
Source: Pure.psu.edu



ICDS Interdisciplinary Initiatives



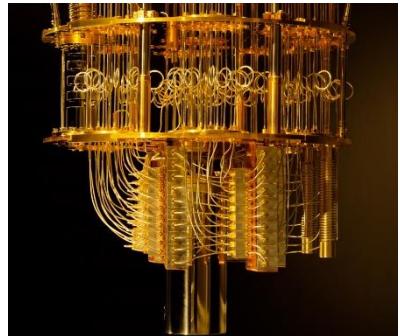
Predictive Science



Infrastructure



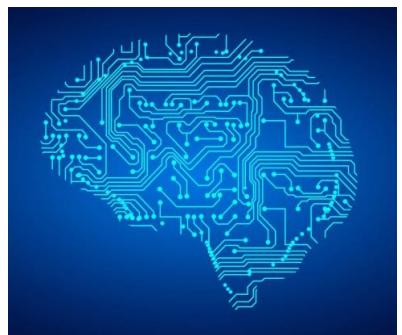
Industry Applications



Quantum Computing



Data Privacy,
Policy and Law



AI/Machine
Learning



Advanced Visualization
& Immersive Experience



Resilience & Social
Justice (Joint)



Diversity in
Computational &
Data Science



International
Partnerships



Example Development



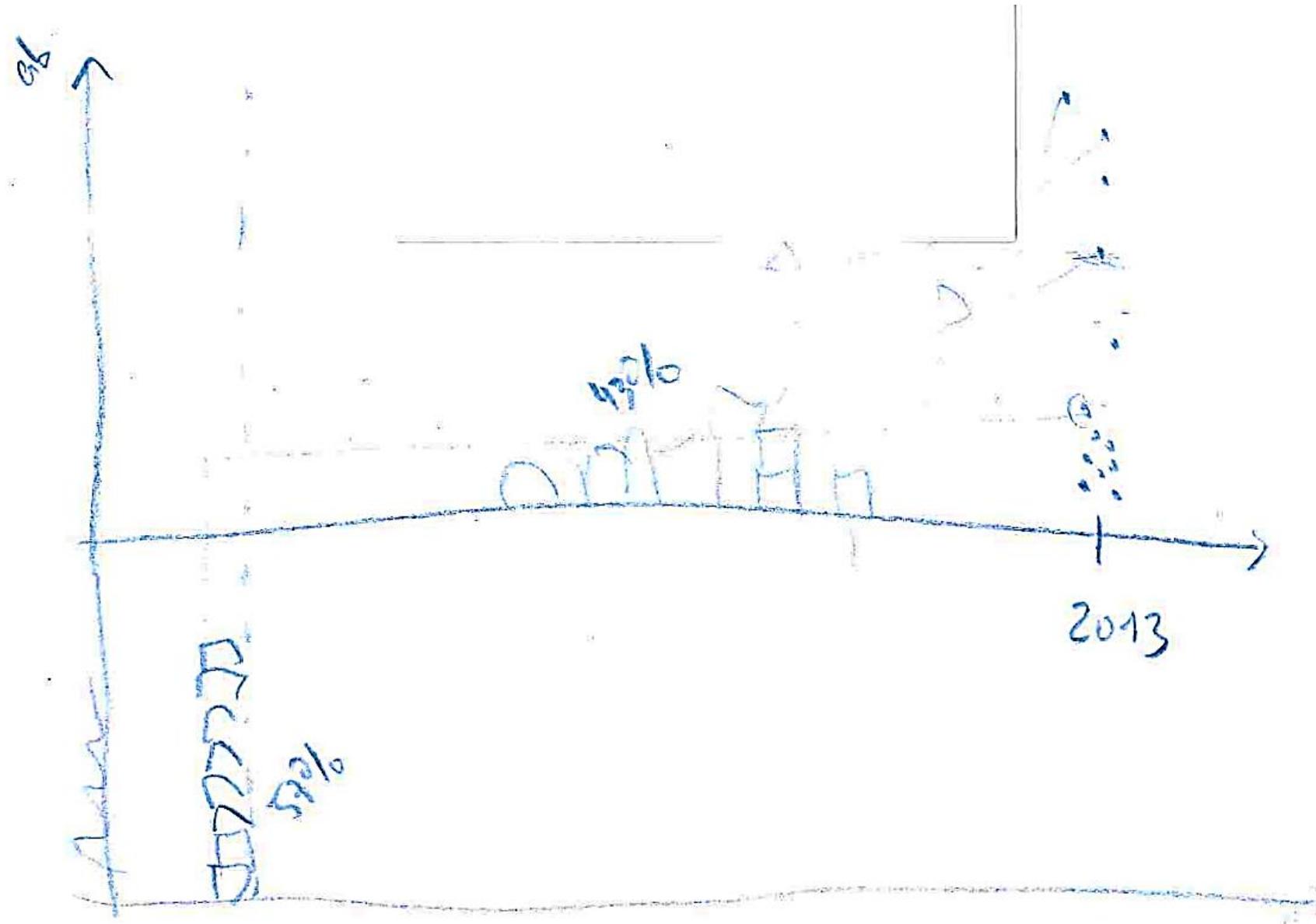
Client

- Dr. Maciej F Boni's academic background is in mathematical epidemiology, mathematical population genetics, individual-based disease transmission models, field epidemiology, phylogenetics, and recombination.
- His main area of focus is the dynamics and evolution of influenza virus, which he has addressed with theoretical, field epidemiological, and informatic methods.
- His main work in malaria focuses on optimal methods of distributing antimalarial drugs to minimize the risks and slow down the spread of drug resistance.



<http://mol.ax/people/>





Sketching is about Design

- Sketching is not about drawing
 - It is about design
- Sketching is a tool to help you:
 - Express,
 - develop, and
 - communicate design ideas
- Sketching is part of a process:
 - idea generation,
 - design elaboration,
 - design choices,
 - engineering





User Analysis

Who will be using/observing the visualizations?

Environment Analysis

Where will this take place?

front-end analysis

Function Analysis

What the visualization provide

Task Analysis

What people need



User Analysis

Who will be using/observing the visualizations?

Characteristics to Consider:

Age

Gender

Education Level

Reading Ability

Physical Size

Physical Abilities/Disabilities

Familiarity to the Product

Task Relevant Skills

- For us:

- Fellow researchers/doctors
- (Age) Roughly 25 - 60
- (Gender) Mixed
- (Familiarity to the Product) Familiar to the concept, but the idea is novel



Environment Analysis

Where will this take place?

Characteristics to Consider:
Location
Lighting
Temperature
Weather
Noise Level
Accessibility

- For us:
 - At a conference
 - During a presentation
 - Wi-Fi?
 - Voice-over



Function Analysis

What the visualization provide

Task Analysis

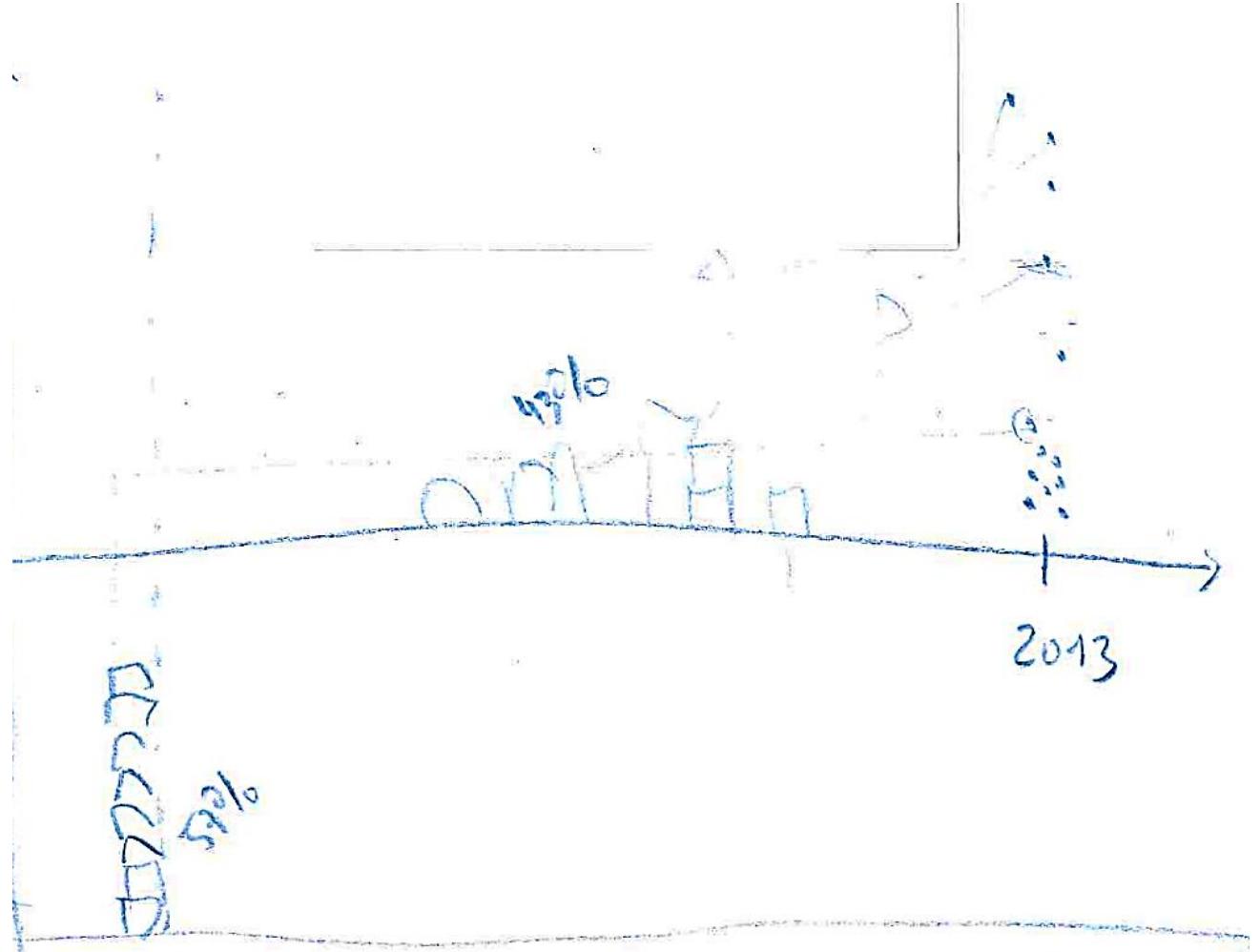
What people need



- What tasks do our users need?
- And
- What are the general functions that your users need to accomplish?
- Task -> Functions



- Task:
 - Present a novel idea that will showcase two specific scenes
- Function:
 - Visualization will need to have a smooth transition between these two scenes
- Task:
 - Show small example to build to a broader idea
- Function:
 - Create an animation that will start with a few points highlighted, then one year, then the entire dataset
- Task:
 - Connect Antibody Count and Date of Infection
- Function:
 - Make a visualization that connects the date of when the antibody count was collected to the actual count. Transition data to show these same data points and when they were infected.



Penn State

Institute for Computational
and Data Sciences

pmd19

Slide:59

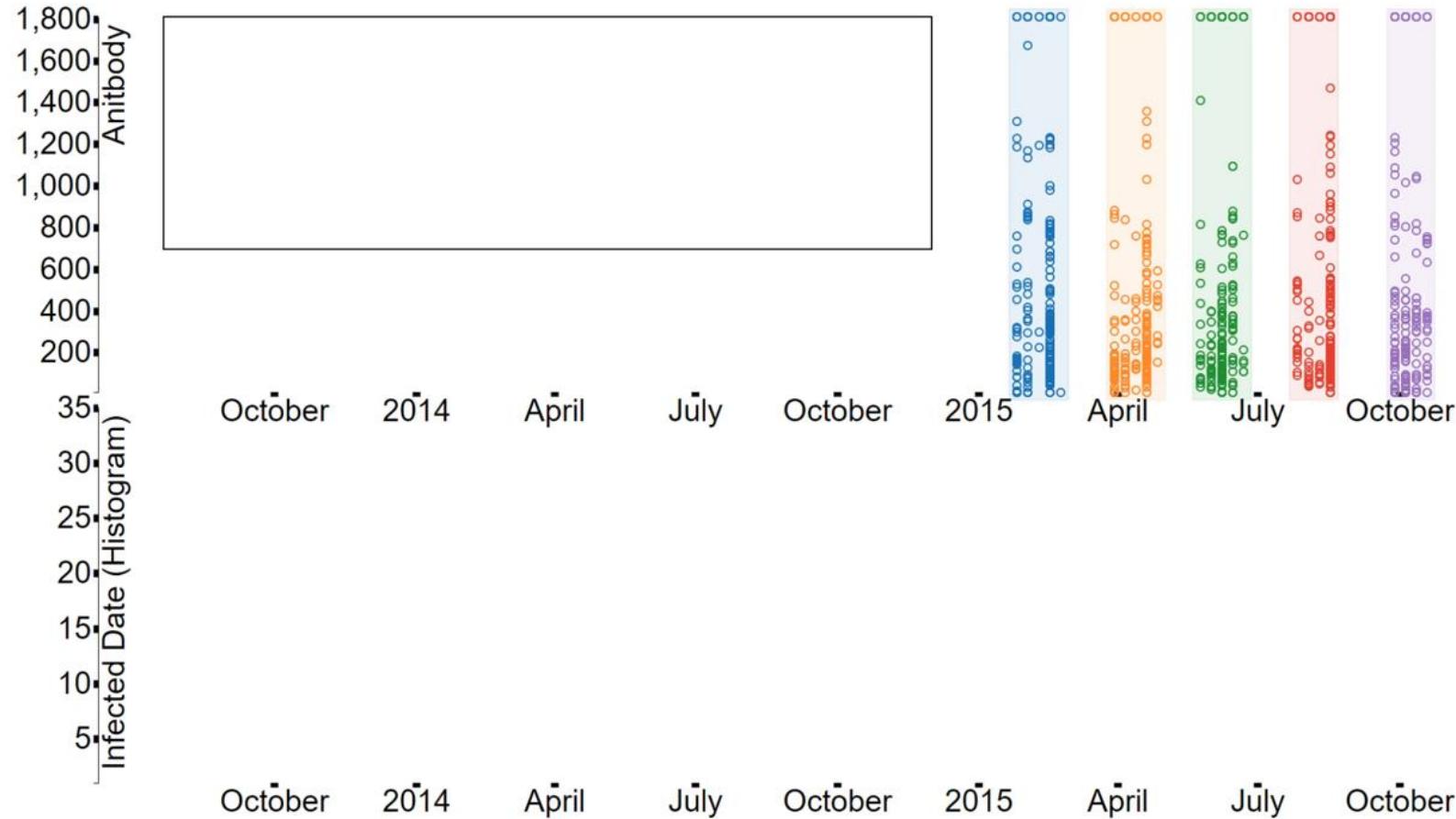


Conclusions

- Design with the intention to create a short video
 - Something that can be embedded into PowerPoint
- Develop a script for transitioning from small dataset to one year to all data



Final Product



**OK, but what about the
Center for Immersive
Experiences?**





PennState

Senior Vice President for Research

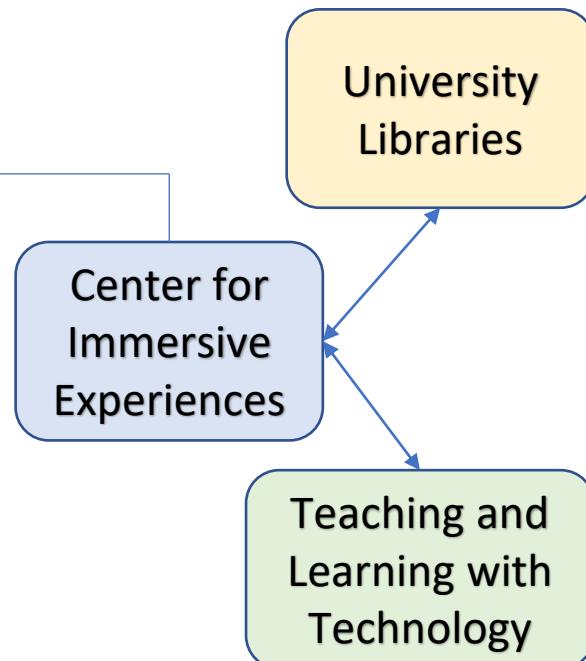
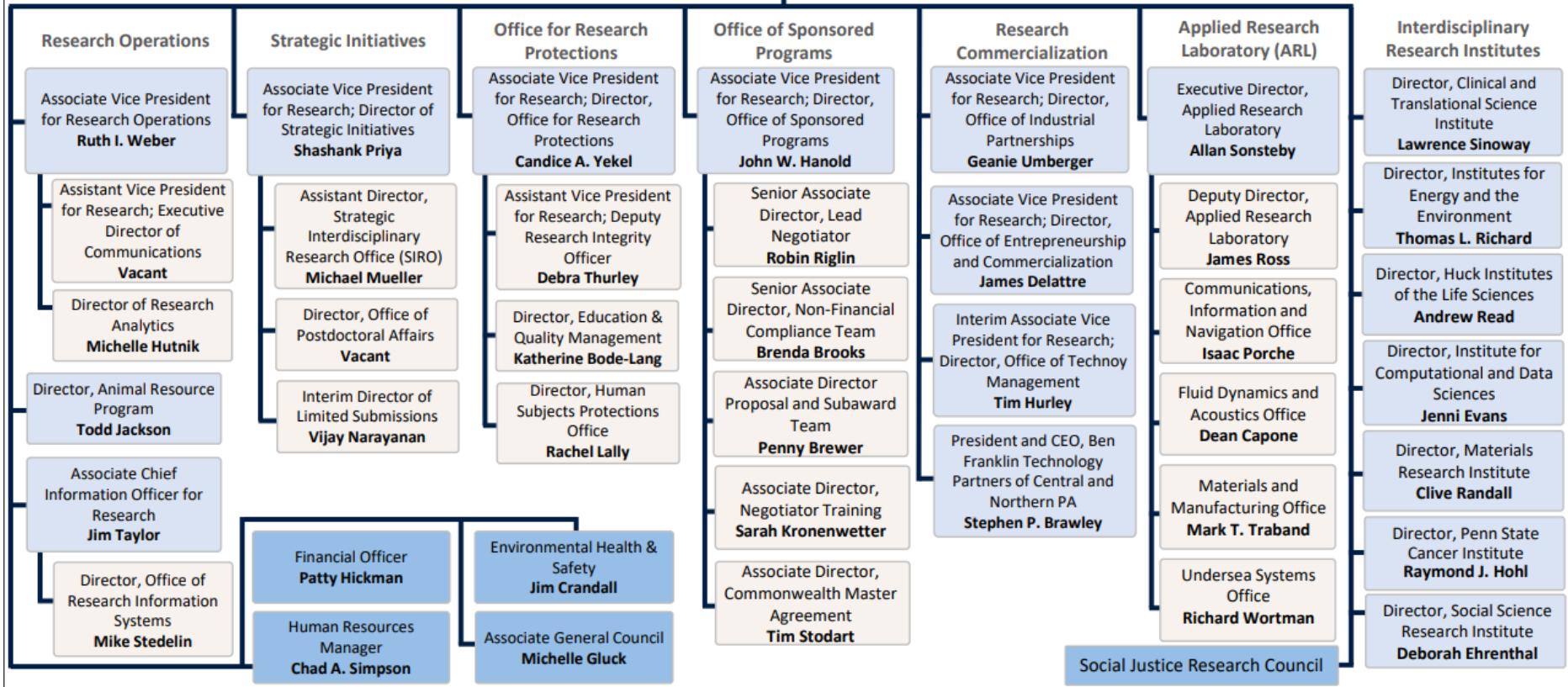
Associate VP Research; Vice Dean for Research and
Graduate Studies, College of Medicine
Leslie J. Parent

Executive Assistant
Diane Plummer

**Senior Vice President
for Research**
Lora G. Weiss

- University Research Council
 - Penn State Research Foundation
 - Research Park Management Corp

Last updated: 1/13/2022



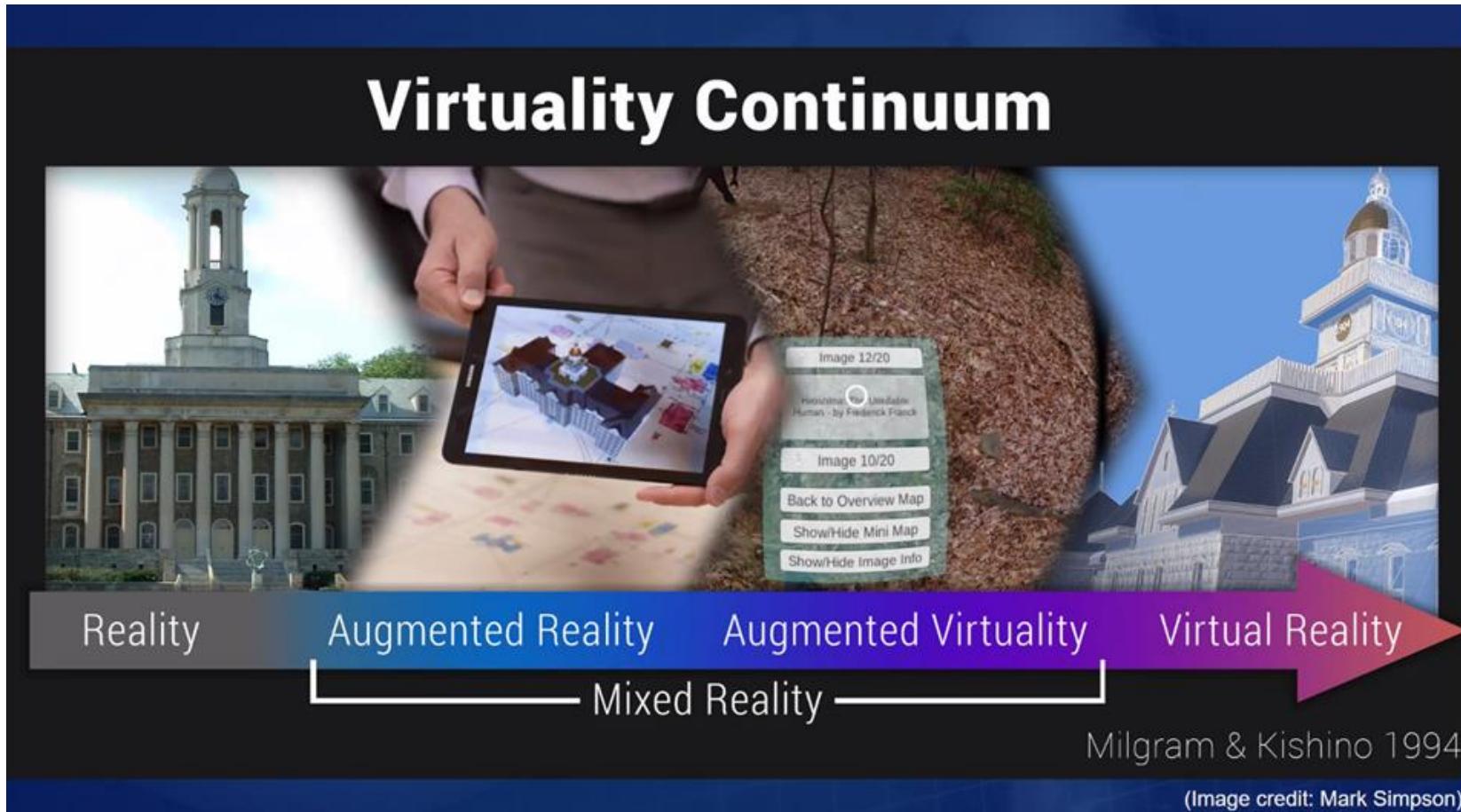
PennState
Institute for Computational
and Data Sciences

pmd19

Slide:63



immersive technologies: virtual, augmented, and mixed reality



the prospects of immersive technologies

- Typically seen in the video game industry
- Facebook bought Oculus for \$2.5 billion (2014)
- Microsoft (HoloLens), Facebook (Oculus), Apple (ARkit 2/shared experiences), Google (VR and AR capability to its Chrome browser/Cardboard), Sony (PlayStation VR)



VR Labs on Campus

The Dreamery – Shields Building (1)

IMEX – Ag Science (2)

Our Location – Center for Immersive Experiences (3)

VR Room – EMS Library (4)



The Team



ZAC ZIDIK



ALEX FATEMI



JAN OLIVER WALLGRÜN



PEJMAN SAJJADI



ERICA KRIEGER



BART MASTERS

<https://immersive.psu.edu/about/meet-the-team/>



Projects on Campus

- Architectural Engineering (iCon)
- ARL (SEALAB)
- Communications (Media Effects Lab)
- IST (Laboratory for Perception, Action, and Cognition)
- Psychology (Brain, Language, and Computation Lab)
- Music Education (Ann Clements)
- Geography (Alex Klippel)
- History (Elizabeth Mansfield)
- Recreation, Park, and Tourism Management (Peter Newman)
- Education (Heather Zimmerman & Susan Land)



So, what is the technology?

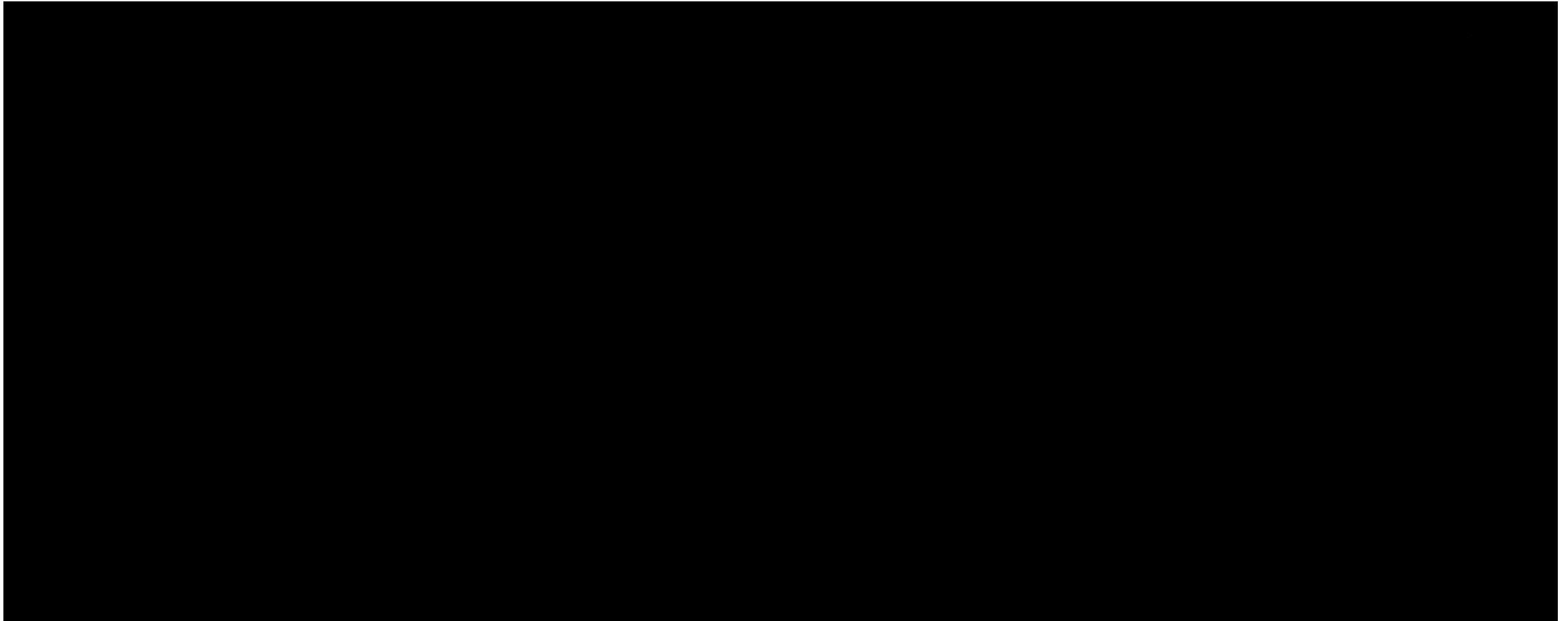


360 Video



Immersive Experiences at Penn State - <https://imex.psu.edu/>





<https://www.facebook.com/PennStateBlueBand/videos/422866724935479>

<https://blueband.psu.edu/media/>



PennState

Institute for Computational
and Data Sciences

pmd19

Slide:71



“Joint”

- Multi-user experiences
- Linking different VR platforms
- Remote & local



- Virtual version of actual field trip
- Traditional classroom / lab setup for higher learning motivation (?)

Designing Low-Cost Joint VR Experiences – Wallgrün et al.



360 Images



- <https://cie.psu.edu/space/>
- <https://cie.psu.edu/space/3Dtour/three.html>



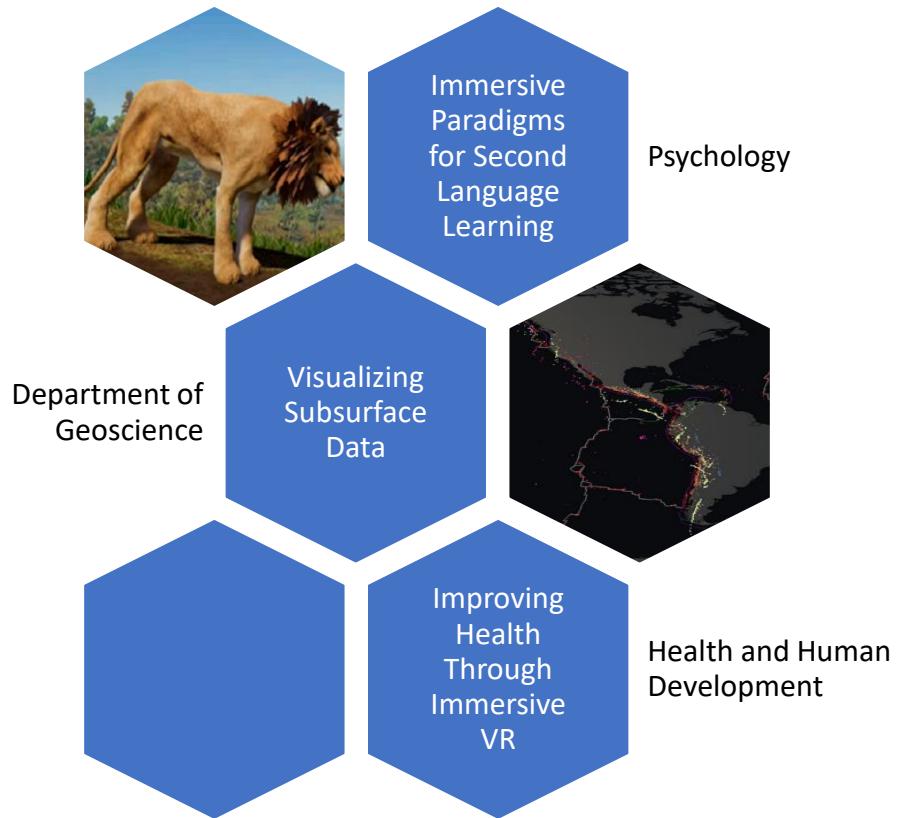
Virtual Reality



Center for Immersive Experiences - <https://immersive.psu.edu/>



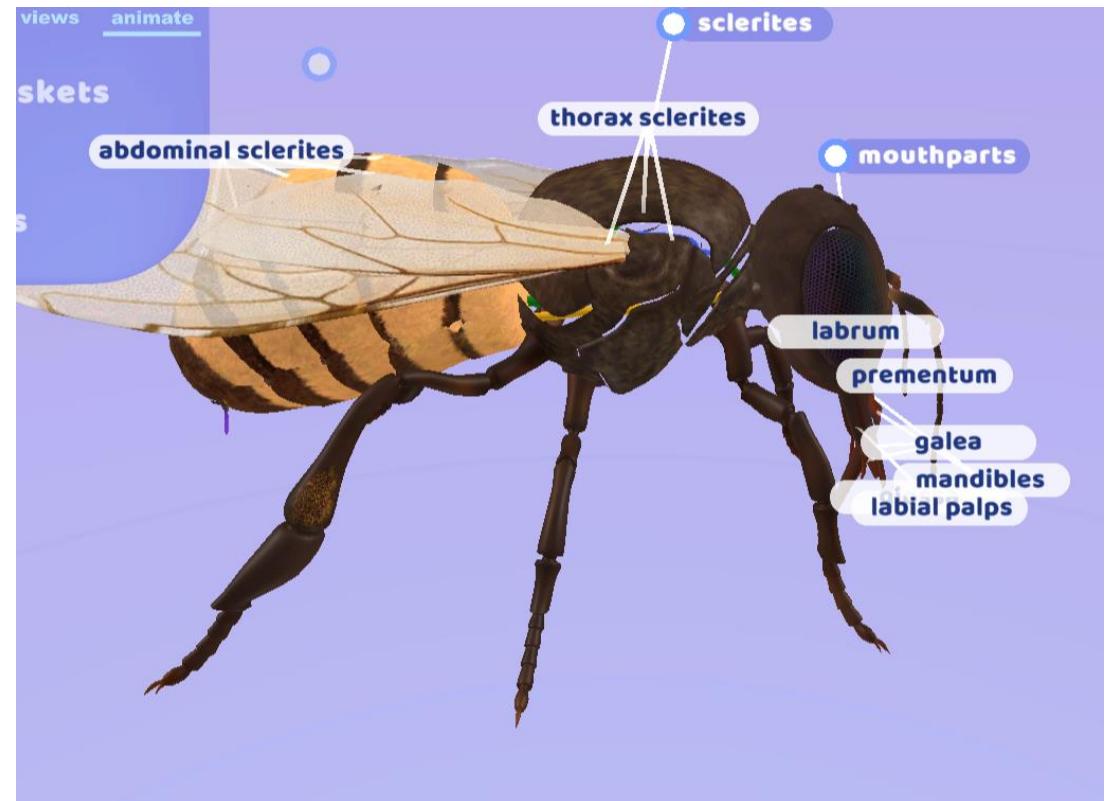
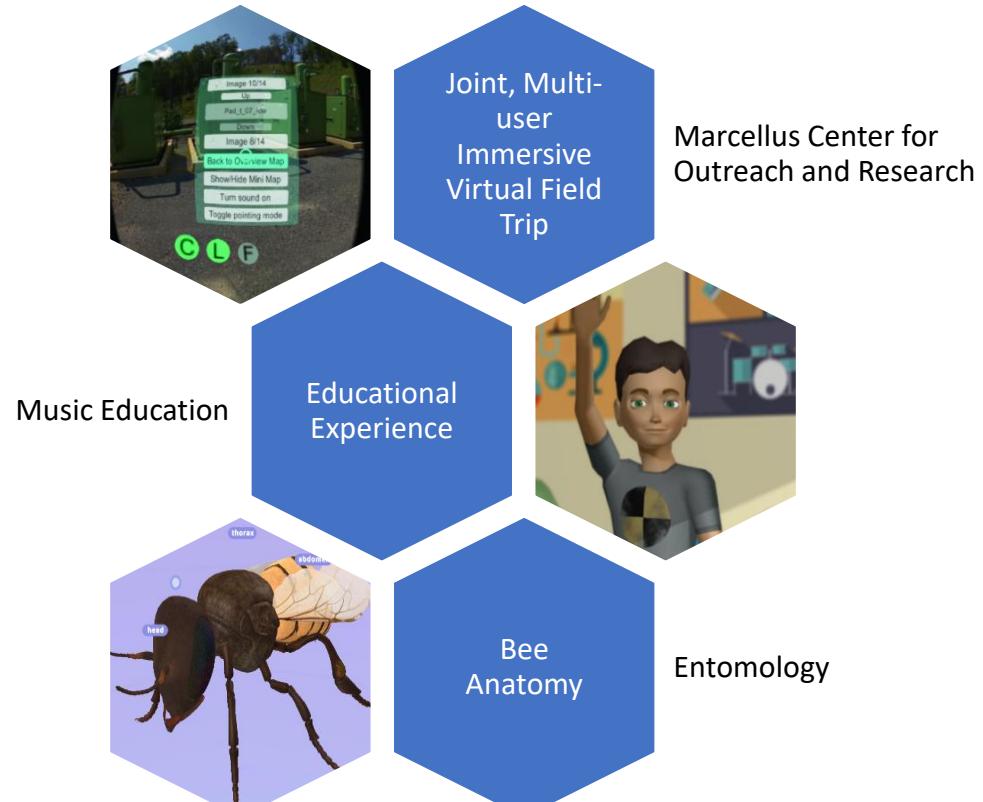
Research Projects



Erica Smithwick - Visualizing Forest Futures Under Climate Uncertainty: Integrating Indigenous Knowledge into Decision-Support Tools for Collaborative Decision Making



Academic Projects/Support



<https://cie.psu.edu/BeeModel/> - Erica Krieger



HoloLens/Augmented Reality



- Adobe Aero



So, what are you interested in?



Accessibility in Immersive Technology



Metaverse



Developing Open Educational Resources



Bridging New Technology to Penn State



Microbiome Week at CIE

- Week of February 21st

