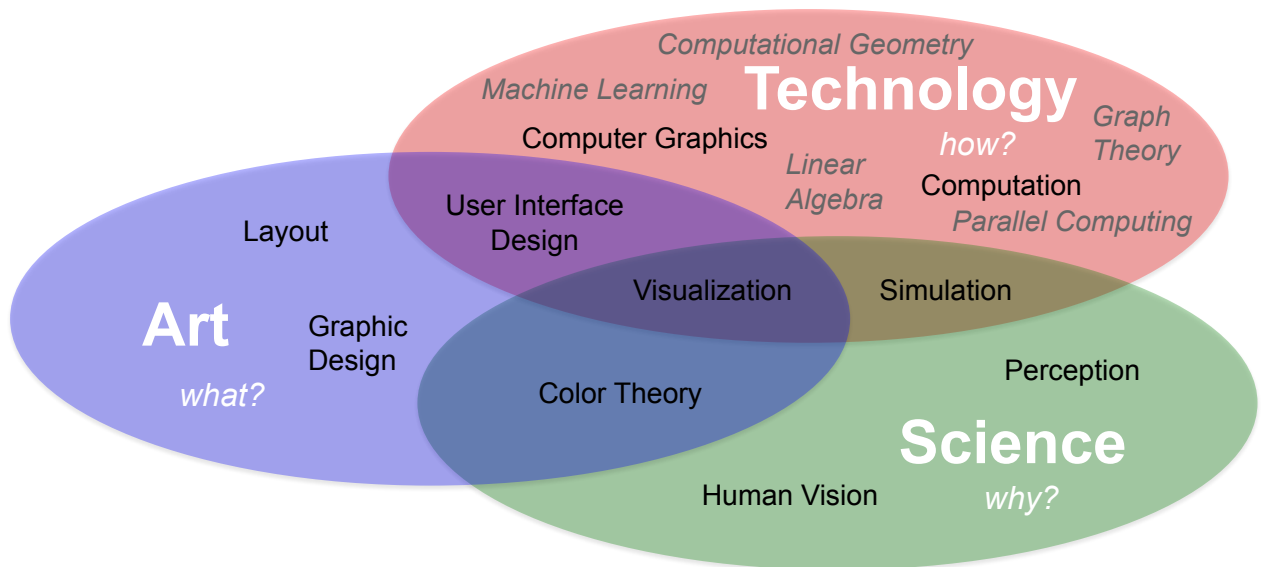


# CSCI 4550/6550 Interactive Visualization

<https://www.cs.rpi.edu/~cutler/classes/visualization/S24/>

## Lecture 1: Introduction to Visualization

### “Introduction” to Visualization



# Today

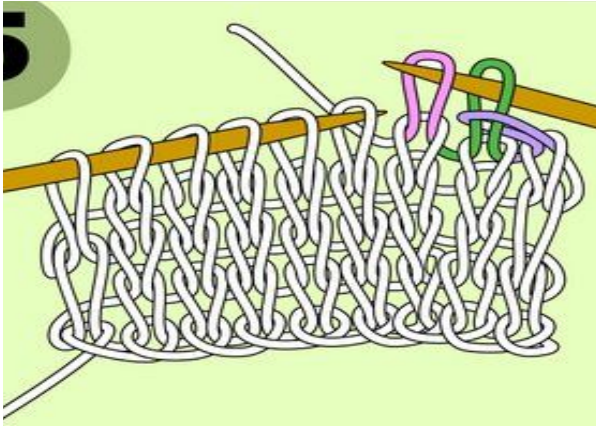
---

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”

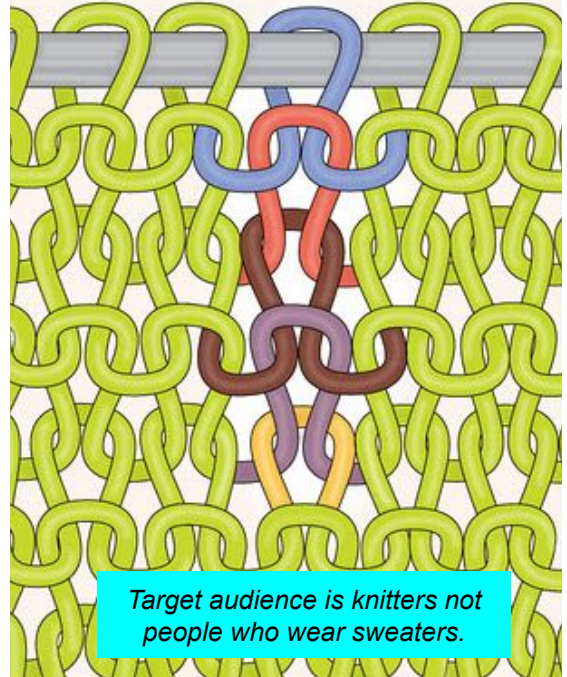
# Today

---

- The Visualization Process
  - Motivation & Problem Definition  
e.g., audience, purpose, goals, interdisciplinary collaboration
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”



<http://techknitting.blogspot.com/>



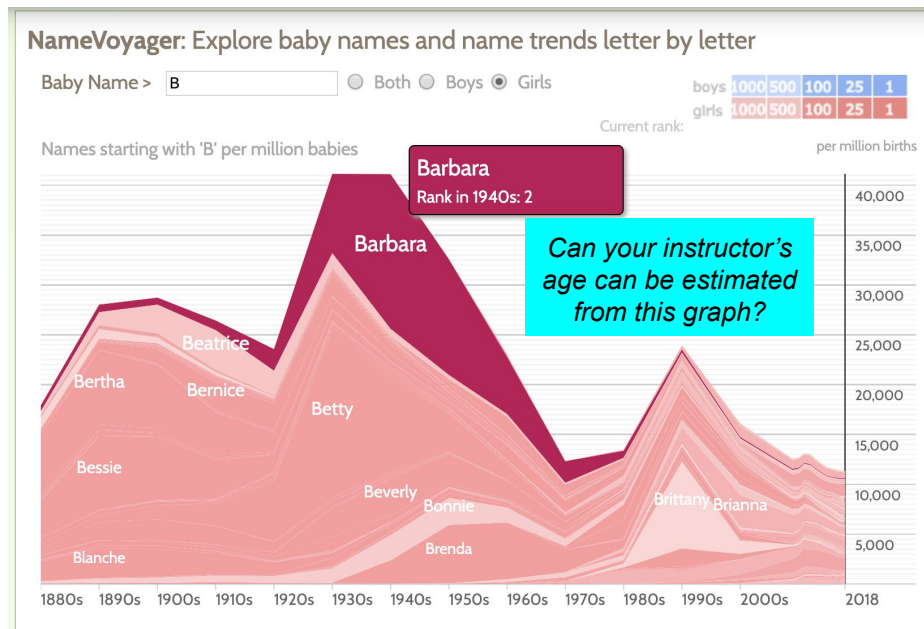
*Target audience is knitters not people who wear sweaters.*

## Today

---

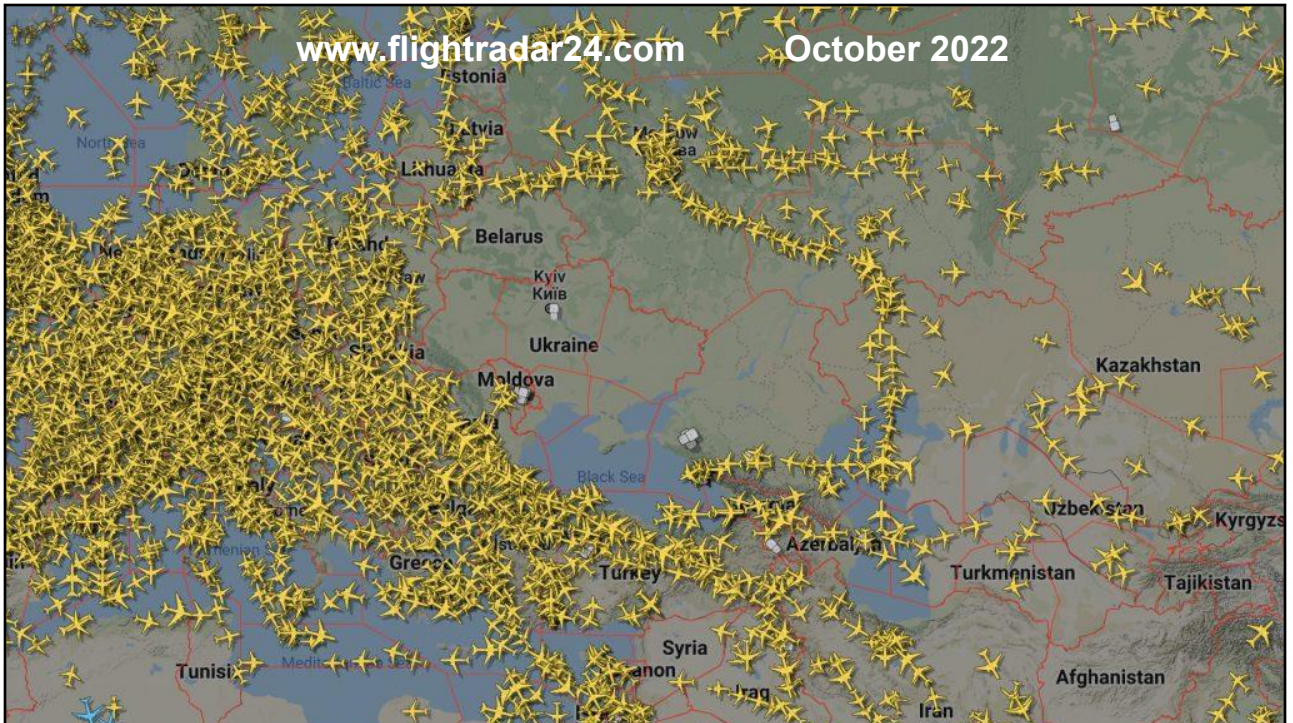
- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
    - e.g., media, color, organization, layout,
    - static vs. dynamic, interactive, art & creativity!
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”

<http://www.babynamewizard.com/voyager>



## Today

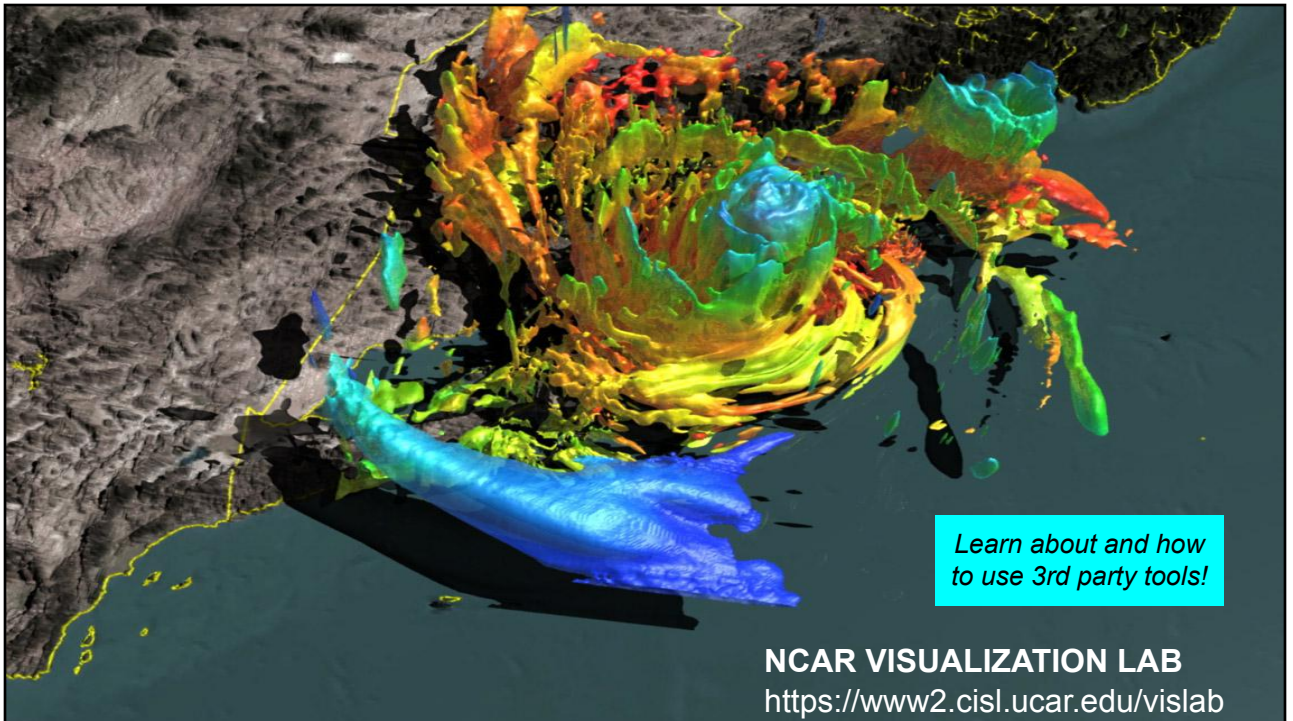
- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
    - e.g., data structures, file formats, parsing, performance & efficiency, databases, very large datasets, interdisciplinary collaboration
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: "Eenie, Meenie, Minie, Moe: Selecting the Right Graph..."



# Today

---

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
    - e.g., data structures, implementation details, visualization toolkits (VTK, OpenGL, d3.js, etc.), performance & efficiency
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”



*Learn about and how  
to use 3rd party tools!*

**NCAR VISUALIZATION LAB**  
<https://www2.cisl.ucar.edu/vislab>

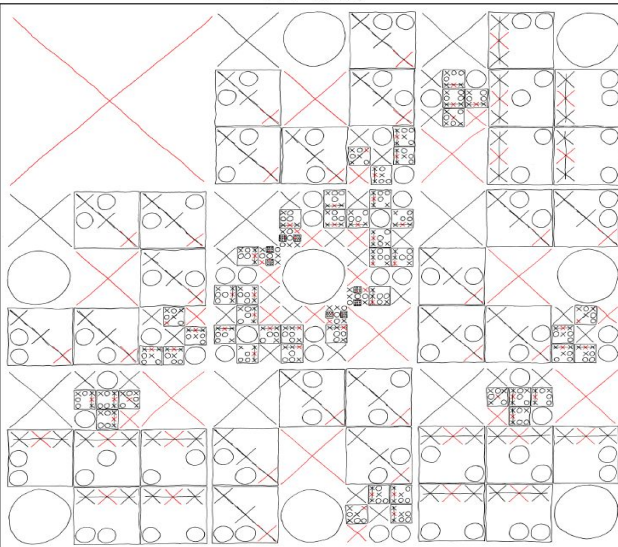
# Today

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
    - e.g., debugging, drawing conclusions from data, accuracy, precision, interpretation, useability
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”

## COMPLETE MAP OF OPTIMAL TIC-TAC-TOE MOVES

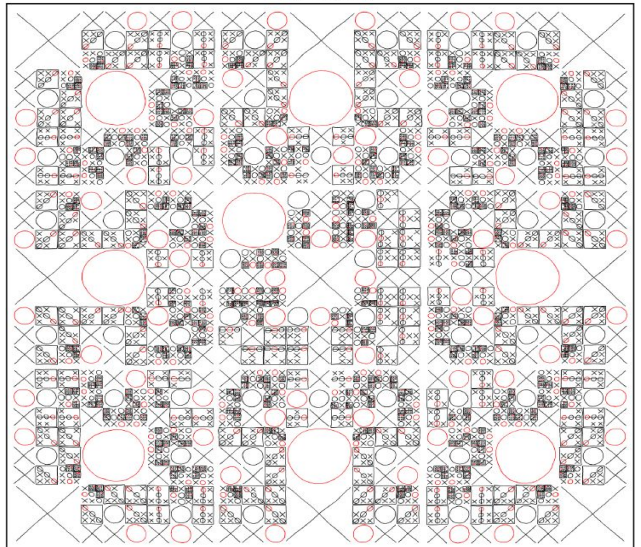
YOUR MOVE IS GIVEN BY THE POSITION OF THE LARGEST RED SYMBOL ON THE GRID. WHEN YOUR OPPONENT PICKS A MOVE, ZOOM IN ON THE REGION OF THE GRID WHERE THEY WENT. REPEAT.

MAP FOR X:



**Make sure the finished visualization is accurate!**

MAP FOR O:



# Today

---

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
    - e.g., prototype & revise, iterated design, comparing before & after, solicit user feedback, formal user studies
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”



<http://www.heraldsun.com.au/Getty>



[http://www.digitalglobe.com/sites/default/files/italy\\_giglio\\_jan17\\_2012\\_0.jpg](http://www.digitalglobe.com/sites/default/files/italy_giglio_jan17_2012_0.jpg)



From somewhere on Facebook....  
<http://www.facebook.com/babayoff>

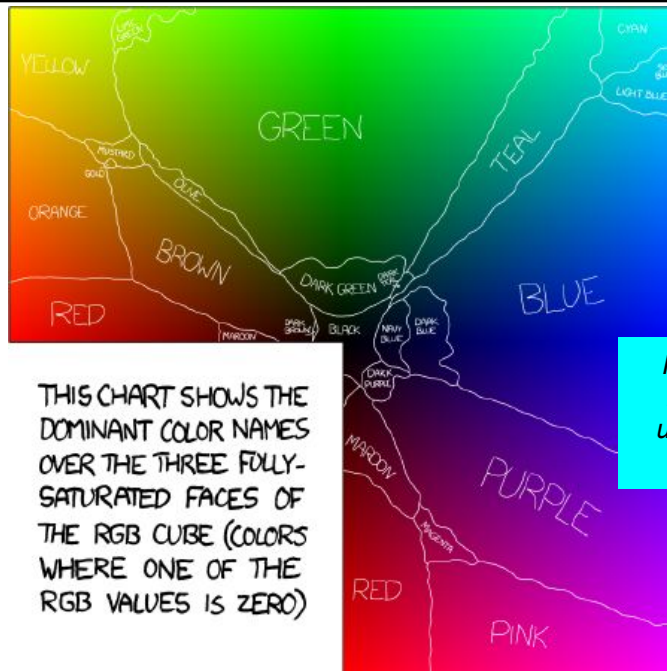
*What is the desired  
takeaway message?*



# Today

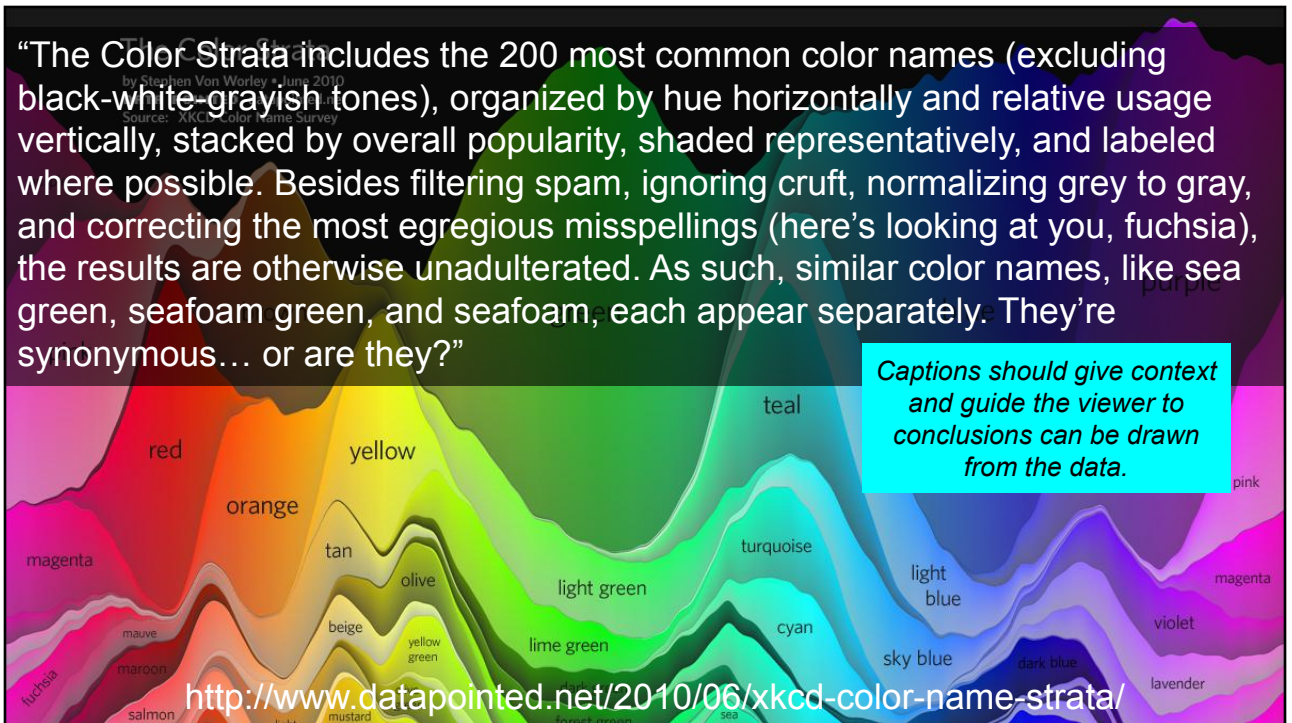
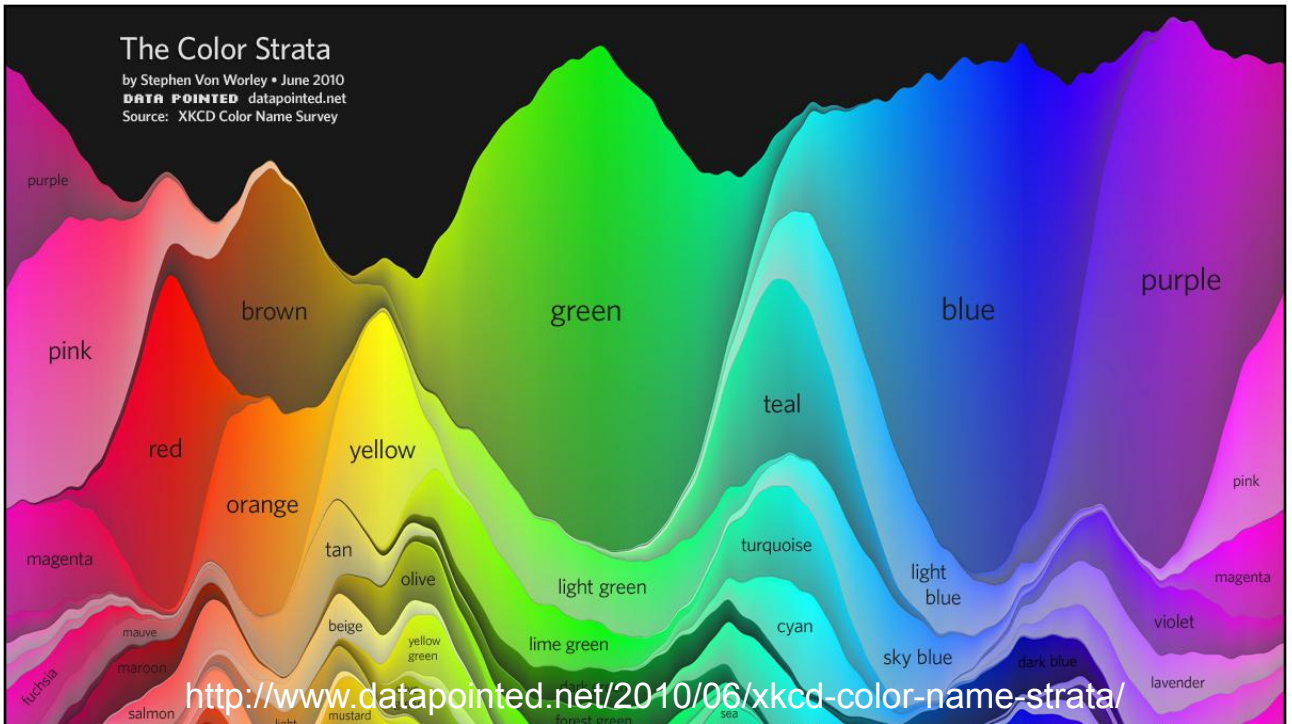
---

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - **Presentation**  
e.g., mixed media, descriptive titles/labels, concise and complete captions/companion text, elevator pitch, documentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”



*It's pretty, but I need more information to understand and draw conclusions.*

<http://blog.xkcd.com/2010/05/03/color-survey-results/>



# Today

---

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”

# Website, Syllabus, etc.

---

RPI Computer Science [↗](#) > [Submittity](#) > [Interactive Visualization](#) [↗](#)



The screenshot shows a navigation menu on the left with the following items: Course Home, Gradeables, Notifications, Discussion Forum, **Syllabus** (highlighted), Grading, Course Calendar, Assigned Readings, Tips for Discussants, and Homework Information. The main content area is titled "Syllabus" and "Course Overview". The "Course Overview" text reads: "Visualizing data is a key step in understanding many problems. This course is designed to introduce students to methods of visualizing many different types of data, such as images, 3D surfaces, flow fields, and medical data. We will both use existing visualization software and program custom visualizations using C++ and OpenGL. Course activities include discussion of recent and classic research papers, weekly homework assignments, in-class critiques of visualization artifacts, and a final project to explore creative uses of these techniques." Below this is the heading "What you should know before taking Interactive Visualization".

<https://www.cs.rpi.edu/~cutler/classes/visualization/S24/syllabus.php>

## “Rules” for the course

---

- As class participation is 5% of your grade:
  - *Using laptops during class is strongly discouraged*
  - If you're using your laptop you need to participate twice as much as everyone else because I'm going to assume you're doing something else.
- Use of laptops for reference during paper discussion is allowed
- *Sit in a different seat, next to different people, each lecture*
  - Work with a different person for each in-class “worksheet”

## Today

---

- The Visualization Process
  - Motivation & Problem Definition
  - Visualization Design
  - Data Collection
  - Visualization Execution
  - Analysis & Validation
  - Visualization Revision
  - Presentation
- Website & Syllabus, Homework 1, etc.
- Reading: “Eenie, Meenie, Minie, Moe: Selecting the Right Graph...”

# Paper for Today & Discussant

---

- "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, 2004
- Normally...
  - Everyone would have made a Submitty Forum post (~200 words) before 10am with a non-trivial *comment* or *question*
  - We'd have one or two student *discussants*... but I'll do that today
  - **Summarize Paper (< 5 min)** *everyone already read the paper!*
  - **Summarize Discussion on Submitty Forum & Moderate In-Class Discussion (~15 min)**

## "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, 2004

---

- **Paper Summary / Contributions (5 mins max!)**
  - Learn conventions/patterns, applies to almost all business data (not necessarily all scientific data)
  - Avoid viewer confusion / mis-information
  - Definition: Categorical (what) vs. Quantitative (how much)
  - Definition: Nominal (order free), Ordinal (ordered), Interval (e.g. histogram)
  - Available visual attributes for encoding data: location, size, shape, orientation, color – *choose wisely!!*
    - Points (scatter plots) & lines (show trends) & bars (emphasize values for categorical data)
    - Size & color less effective for encoding quantitative values

Type/Description	Encoding Methods	Example																										
<b>Nominal Comparison</b> A simple comparison of the categorical subdivisions of one or more measures in no particular order	<ul style="list-style-type: none"> <li>Bars only (horizontal or vertical)</li> </ul>	<p>Q1 2003 Calls by Region</p> <table border="1"> <tr><th>Region</th><th>Calls</th></tr> <tr><td>North</td><td>3,000</td></tr> <tr><td>East</td><td>4,500</td></tr> <tr><td>South</td><td>2,500</td></tr> <tr><td>West</td><td>5,000</td></tr> </table>	Region	Calls	North	3,000	East	4,500	South	2,500	West	5,000																
Region	Calls																											
North	3,000																											
East	4,500																											
South	2,500																											
West	5,000																											
<b>Time Series</b> Multiple instances of one or more measures taken at equidistant points in time	<ul style="list-style-type: none"> <li>Lines to emphasize overall pattern</li> <li>Bars to emphasize individual values</li> <li>Points connected by lines to slightly emphasize individual values while still highlighting the overall pattern</li> <li>Always place time on the horizontal axis</li> </ul>	<p>2003 Sales</p> <table border="1"> <tr><th>Month</th><th>Sales</th></tr> <tr><td>Jan</td><td>2,500</td></tr> <tr><td>Feb</td><td>2,800</td></tr> <tr><td>Mar</td><td>2,500</td></tr> <tr><td>Apr</td><td>3,000</td></tr> <tr><td>May</td><td>2,800</td></tr> <tr><td>Jun</td><td>3,200</td></tr> <tr><td>Jul</td><td>2,800</td></tr> <tr><td>Aug</td><td>3,000</td></tr> <tr><td>Sep</td><td>3,200</td></tr> <tr><td>Oct</td><td>3,500</td></tr> <tr><td>Nov</td><td>3,800</td></tr> <tr><td>Dec</td><td>4,000</td></tr> </table>	Month	Sales	Jan	2,500	Feb	2,800	Mar	2,500	Apr	3,000	May	2,800	Jun	3,200	Jul	2,800	Aug	3,000	Sep	3,200	Oct	3,500	Nov	3,800	Dec	4,000
Month	Sales																											
Jan	2,500																											
Feb	2,800																											
Mar	2,500																											
Apr	3,000																											
May	2,800																											
Jun	3,200																											
Jul	2,800																											
Aug	3,000																											
Sep	3,200																											
Oct	3,500																											
Nov	3,800																											
Dec	4,000																											
<b>Ranking</b> Categorical subdivisions of a measure ordered by size (either descending or ascending)	<ul style="list-style-type: none"> <li>Bars only (horizontal or vertical)</li> <li>To highlight high values, sort in descending order</li> <li>To highlight low values, sort in ascending order</li> </ul>	<p>Headcount</p> <table border="1"> <tr><th>Department</th><th>Headcount</th></tr> <tr><td>Manufacturing</td><td>250</td></tr> <tr><td>Sales</td><td>180</td></tr> <tr><td>Engineering</td><td>100</td></tr> <tr><td>Operations</td><td>80</td></tr> <tr><td>Finance</td><td>60</td></tr> <tr><td>Info Systems</td><td>50</td></tr> <tr><td>Legal</td><td>30</td></tr> <tr><td>Marketing</td><td>20</td></tr> </table>	Department	Headcount	Manufacturing	250	Sales	180	Engineering	100	Operations	80	Finance	60	Info Systems	50	Legal	30	Marketing	20								
Department	Headcount																											
Manufacturing	250																											
Sales	180																											
Engineering	100																											
Operations	80																											
Finance	60																											
Info Systems	50																											
Legal	30																											
Marketing	20																											
<b>Part-to-Whole</b> Measures of individual categorical subdivisions as ratios to the whole	<ul style="list-style-type: none"> <li>Bars only (horizontal or vertical)</li> <li>Use stacked bars only when you must display measures of the whole as well as the parts</li> </ul>	<p>Regional % of Total Expenses</p> <table border="1"> <tr><th>Region</th><th>% of Total Expenses</th></tr> <tr><td>West</td><td>35%</td></tr> <tr><td>East</td><td>30%</td></tr> <tr><td>North</td><td>20%</td></tr> <tr><td>South</td><td>15%</td></tr> </table>	Region	% of Total Expenses	West	35%	East	30%	North	20%	South	15%																
Region	% of Total Expenses																											
West	35%																											
East	30%																											
North	20%																											
South	15%																											

"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message",  
 Stephen Few, 2004

<b>Deviation</b> Categorical subdivisions of a measure compared to a reference measure, expressed as the differences between them	<ul style="list-style-type: none"> <li>Lines to emphasize the overall pattern only when displaying deviation and time-series relationships together</li> <li>Points connected by lines to slightly emphasize individual data points while also highlighting the overall pattern when displaying deviation and time-series relationships together</li> <li>Bars to emphasize individual values, but limit to vertical bars when a time-series relationship is included</li> <li>Always include a reference line to compare the measures of deviation against</li> </ul>	<p>Actual to Plan Variance</p> <table border="1"> <tr><th>Department</th><th>Variance</th></tr> <tr><td>Sales</td><td>+8</td></tr> <tr><td>Marketing</td><td>-6</td></tr> <tr><td>Systems</td><td>-2</td></tr> <tr><td>HR</td><td>-18</td></tr> <tr><td>Finance</td><td>+5</td></tr> </table>	Department	Variance	Sales	+8	Marketing	-6	Systems	-2	HR	-18	Finance	+5		
Department	Variance															
Sales	+8															
Marketing	-6															
Systems	-2															
HR	-18															
Finance	+5															
<b>Frequency Distribution</b> Counts of something per categorical subdivisions (intervals) of a quantitative range	<ul style="list-style-type: none"> <li>Vertical bars to emphasize individual values (called a <i>histogram</i>)</li> <li>Lines to emphasize the overall pattern (called a <i>frequency polygon</i>)</li> </ul>	<p>Order Count by Order Size</p> <table border="1"> <tr><th>Order Size Range</th><th>Order Count</th></tr> <tr><td>&lt; \$10</td><td>5,000</td></tr> <tr><td>&gt;= \$10 &amp; &lt; \$50</td><td>1,500</td></tr> <tr><td>&gt;= \$50 &amp; &lt; \$100</td><td>1,000</td></tr> <tr><td>&gt;= \$100 &amp; &lt; \$200</td><td>500</td></tr> <tr><td>&gt;= \$200 &amp; &lt; \$400</td><td>300</td></tr> <tr><td>&gt;= \$400 &amp; &lt; \$500</td><td>100</td></tr> </table>	Order Size Range	Order Count	< \$10	5,000	>= \$10 & < \$50	1,500	>= \$50 & < \$100	1,000	>= \$100 & < \$200	500	>= \$200 & < \$400	300	>= \$400 & < \$500	100
Order Size Range	Order Count															
< \$10	5,000															
>= \$10 & < \$50	1,500															
>= \$50 & < \$100	1,000															
>= \$100 & < \$200	500															
>= \$200 & < \$400	300															
>= \$400 & < \$500	100															
<b>Correlation</b> Comparisons of two paired sets of measures to determine if as one set goes up the other set goes either up or down in a corresponding manner, and if so, how strongly	<ul style="list-style-type: none"> <li>Points and a trend line in the form of a scatter plot</li> <li>Bars may be used, arranged as a <i>paired bar graph</i> or a <i>correlation bar graph</i>, if scatter plots are unfamiliar</li> <li>(Note: For descriptions of these graphs, see my book <i>Show Me the Numbers</i>.)</li> </ul>	<p>Correlation of Employee Heights and Salaries</p>														

"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message",  
 Stephen Few, 2004

# "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, 2004

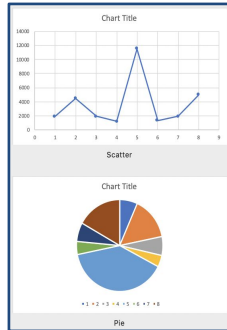
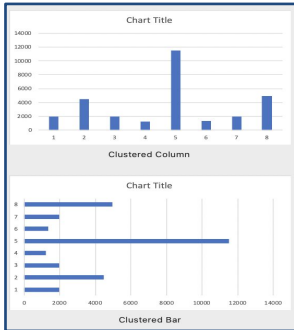
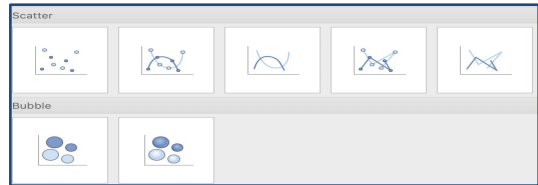
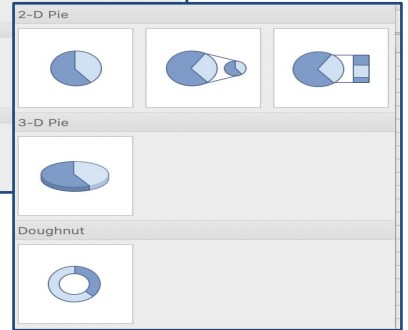
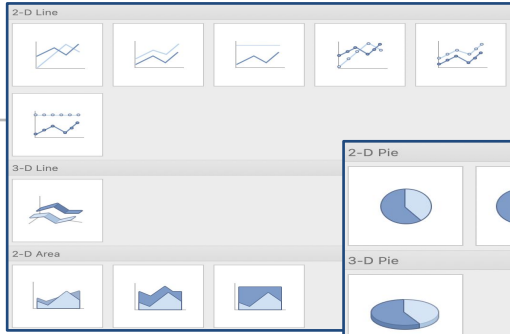
- **Key Points from Submitty Discussion (~15+ mins)**

- This is a nice reference / cheat sheet
- Information was obvious / intuitive
- These are all the charts you can make in Excel / Google Sheets
- Including "bad" examples might have been educational
- What about creativity? (Do we always have to follow the rules?)
- (Un)intentional data misrepresentation... *Ethics are important!*
- What about pie charts? What about spatial/map data?
- Are there similar rules for 3D, interaction, color, etc??
- What about non-scientific data? "information visualization"
- Writing / Communication: This is *not* a normal research paper (e.g., missing references to other papers)

## Google Sheets



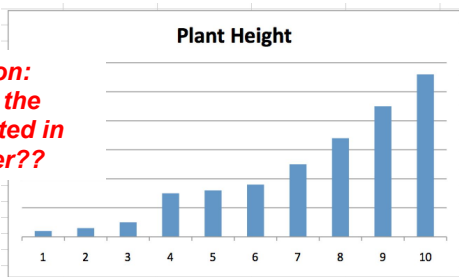
# Microsoft Excel



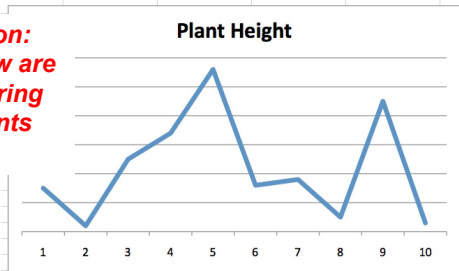
## What I learned in 7<sup>th</sup> grade Science Fair:

1. Always label your axes!
2. Choose the correct chart!

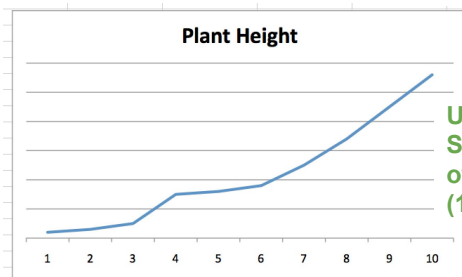
**Confusion:**  
Why are the bars sorted in this order??



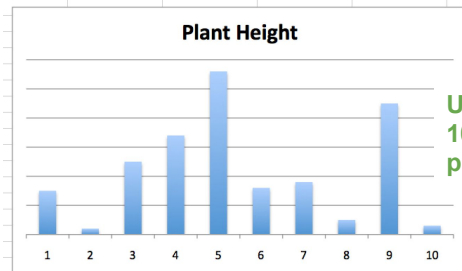
**Confusion:**  
Why/How are neighboring data points linked?



**Use Case:**  
Same plant over time (10 days)



**Use Case:**  
10 different plants





# How *not* to sort Ordinal Data

- It's mid-April... how much \$ is remaining in the travel budget?

April '18	0
August '18	1200
December '18	1950
February '19	1950
January '19	1350
July '18	0
June '18	0
March '19	4950
May '18	0
November '18	1950
October '18	4450
September '18	1950
Expenditures	19750
Total Grant	25000

**Confusion:**  
*Don't sort months alphabetically!*

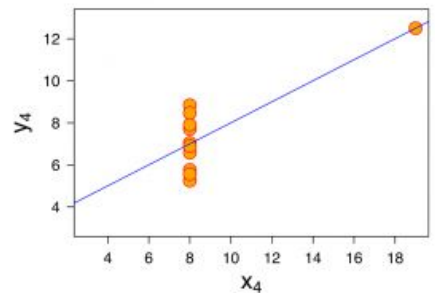
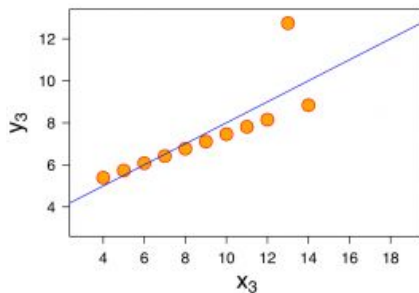
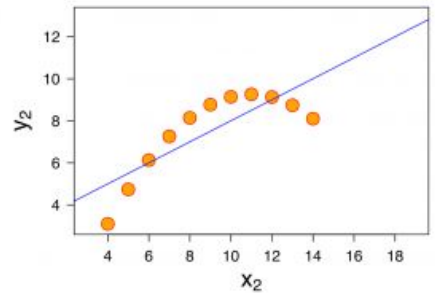
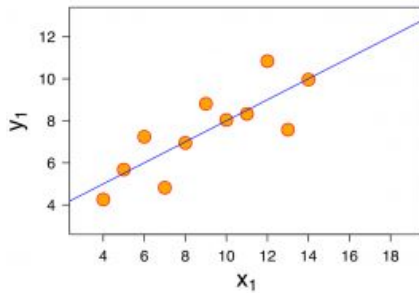
**Proper chronological order is actually useful**

	student pay	travel	total
July '18	0		0
August '18	1200		1950
September '18	1950		1950
October '18	1950	2500	4450
November '18	1950		1950
December '18	1950		1200
<b>Fall 2018 Total</b>	<b>9000</b>	<b>2500</b>	<b>11500</b>
January '19	1350		1350
February '19	1950		1950
March '19	1950	3000	4950
April '19	?		
May '19	?		
June '19	?	?	
<b>Spring 2019 Total</b>	<b>5250</b>	<b>3000</b>	<b>8250</b>
<b>Remaining Budget</b>	<b>3750</b>	<b>1500</b>	<b>5250</b>
<b>Total</b>	<b>18000</b>	<b>7000</b>	<b>25000</b>

These 4 data sets are quite different

But coincidentally all have the same:

- mean,
- variance,
- correlation,
- and
- regression



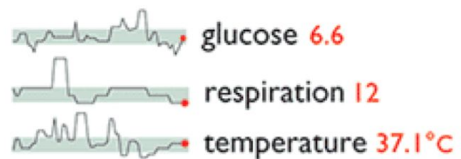
Francis Anscombe [http://en.wikipedia.org/wiki/File:Anscombe%27s\\_quartet\\_3.svg](http://en.wikipedia.org/wiki/File:Anscombe%27s_quartet_3.svg)

# Sparklines – intense *word-size* graphics

- Term coined by Edward Tufte in book *Beautiful Evidence*
- About the size of a word
- Can be embedded within a paragraph of text

The Phillies 2017 season as a sparkline: . Compare that to the 2017 Washington Nationals season: . The one element missing from my implementation is an indicator for home games.

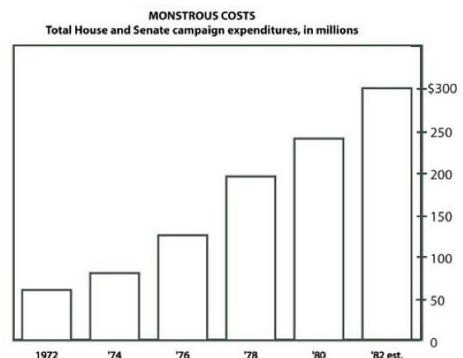
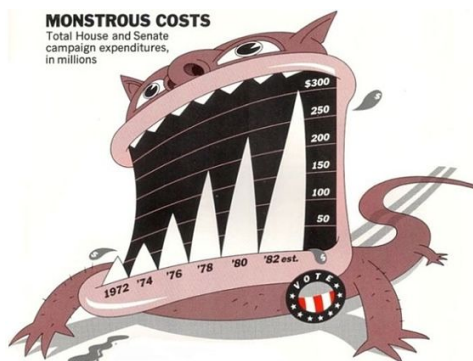
- Quantified
- Variation over time
- Highlight current/last measurement
- Range of what's normal



## Reading for Friday

*Post a comment/question on Submittly Discussion Forum by Friday @10am*

- “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”, Bateman et al., CHI 2010.



- Article discussed here: <http://eagereyes.org/criticism/chart-junk-considered-useful-after-all>

# Homework 1: Inspirational Visualization Images

---

- Find two example visualization *images*:
  - one great visualization
  - one example that needs revision to be effective
- For each example write a paragraph or two describing:
  - the author (full bibliographic citation), context, audience, original media format and purpose of the visualization
  - your analysis of the positive and negative aspects of each example and how it could be improved, and
  - any generalizations you can make about what makes for a compelling, high-quality visualization
- Upload your assignment to Submittly by **11:59pm on Thursday**. And post one of the images on the Submittly discussion forum.

## Is this a Visualization?

---

“Been wondering for years where it is cats put their feet when they settle down into this pose”

“Whoa, so that’s how they do it!”



<https://www.deviantart.com/theninthwavetnw/art/Upo-375998592>

## Criteria for label “(good) Visualization”

---

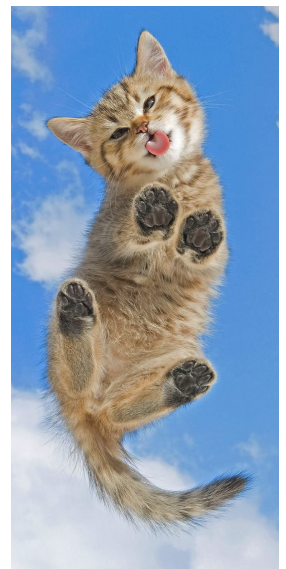
- 

## Criteria for label “(good) Visualization”

---

*From Spring 2020*

- Needs to have numbers/be based on data -- or maybe not?
  - Not all visualization have quantitative data, coordinates
  - Flow charts are visualization
- Puts an image to something
- Should not have extra stuff, should be simplified to show the point/purpose, don't show unnecessary context
- Not just an observation, should be an abstraction of the information
- Adds something, serves a purpose (augmented or cross section),
- a table of data isn't a visualization, but a graph of data is
- Convey information by showing a view that is not normally seen
- If this isn't, what is???
- If you didn't work hard to produce it, it is not a visualization
- Need data to make comparison, help make conclusion/model
- Should have companion text/explanation, should be labeled
- Good use of image, hook to get you to read the text
- Would be better if it were animated, multi-frame



## Criteria for label “(good) Visualization”

From Fall 2018

- Some image, with text, describes purpose
  - Maybe animation/interaction can substitute for needed text
  - Or maybe not even text required (could rely on context/convention/intuition/human experience)
  - Text should be concise
  - Maybe pictogram instead of text (symbols for good vs bad)
- Clearly convey intended information, be clear
  - If text is needed... its there, if not then don't
- Address question or concern - Purpose should be clear
- Be accurate, don't misrepresent data
- Has added value: more intuitive than text (a reason why not just text), information that is not easily available (or ok if its a shift in perspective a new way of looking at info)
- Fit into the argument of paper (the context)
- Understand the audience
- Pleasing to look at
- Avoid superfluous information
- Should cite sources! (available, but off to the side)
  - Give credit to author - Verify the credibility



Under-Dogs: I Photograph Dogs from Underneath, Andrius Burba.  
<https://www.boredpanda.com/under-dogs-i-photograph-dogs-from-underneath/>

## Criteria for label “(good) Visualization”

From Spring 2016

- Author choice
- Not innately visual, the author transformed it to be visual
- Clearly (perfectly uncluttered) data
- Defined metrics (science)
- Highlight important aspects of the dataset
- Intention/purpose?
- Need more than one datapoint, need to show a trend, want to generalize? Do we need time? Maybe not.
  - Comparisons can be very valuable. Sufficient quantity of data to draw conclusions. Other similar datapoints, or datapoints for comparison.
- Needs to be interactive! Want to move the cat around, virtual reality cat!



# Criteria for label “(good) Visualization”

From Fall 2014

- Reveals something you didn't know (about cats)
- Needs to have an X & Y axis (not really?) instead... Needs to exist in a space
- Should be clear in meaning & purpose
- Just enough information and no more
- Intentional
- If the thing can be quantified, must indicate precision & accuracy
- Shouldn't be trying to mislead you but ok to have busy-ness to express the complexity
- Be a scientist, have a hypothesis but look at the data with fresh eyes – don't bias your conclusions, allow for interpretation



[https://www.reddit.com/r/pics/comments/7srqkj/this\\_is\\_why\\_turtle\\_cannot\\_come\\_out\\_from\\_its\\_shell/](https://www.reddit.com/r/pics/comments/7srqkj/this_is_why_turtle_cannot_come_out_from_its_shell/)



<https://imgur.com/gallery/BuyO47o>