

# Data Visualization In A Nutshell

*A brief to the beauty of visualization and its design*

Zhenyuan Lu, [zhenyuanlu.com](http://zhenyuanlu.com)

Zhenyuan Lu

Lived, worked and studied in US for about 9 years

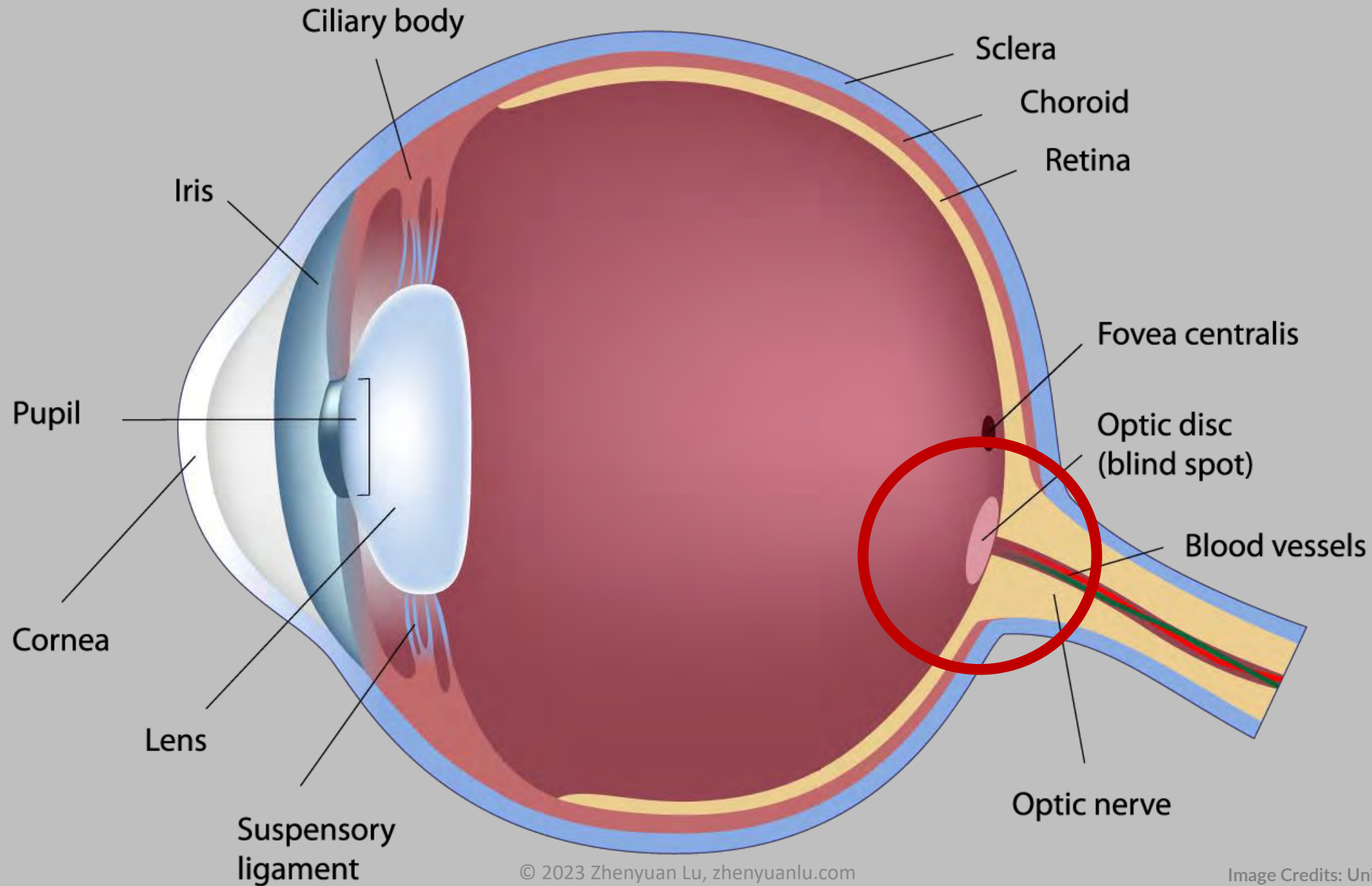
Teaching IE6600,

Course webpage: [zhenyuanlu.com/ie6600-vbc-fa22/](http://zhenyuanlu.com/ie6600-vbc-fa22/)

 This guy is me



Cover your right eye with your left hand and focus your left eye on the cross. Now, with your left eye fixed on the cross, move the paper very slowly closer to your face.



Cognition

# Visualization

# Cognition



On the left, you see what your brain thinks you are seeing. On the right, you see what your eyes are really getting: tons of quick, narrow fixations, such as these views.

## Content

1. What is Data Viz ?
2. Why Data Viz ?
3. How to Build Data Viz ?
4. Real World Issues
5. Q&A

# What is Data Visualization



The main goal of data visualization is to communicate information clearly and efficiently, through graphical means.

- Vitaly Friedman (2008)

Data visualization is a combination of numerous fields whose purpose is to communicate information visually with a specific goal to a specific target audience.

Decision  
Information  
Analysis  
Explore  
Modeling

...

Statistics  
Computer Science  
Information Design  
Operation Research  
Business Management

...

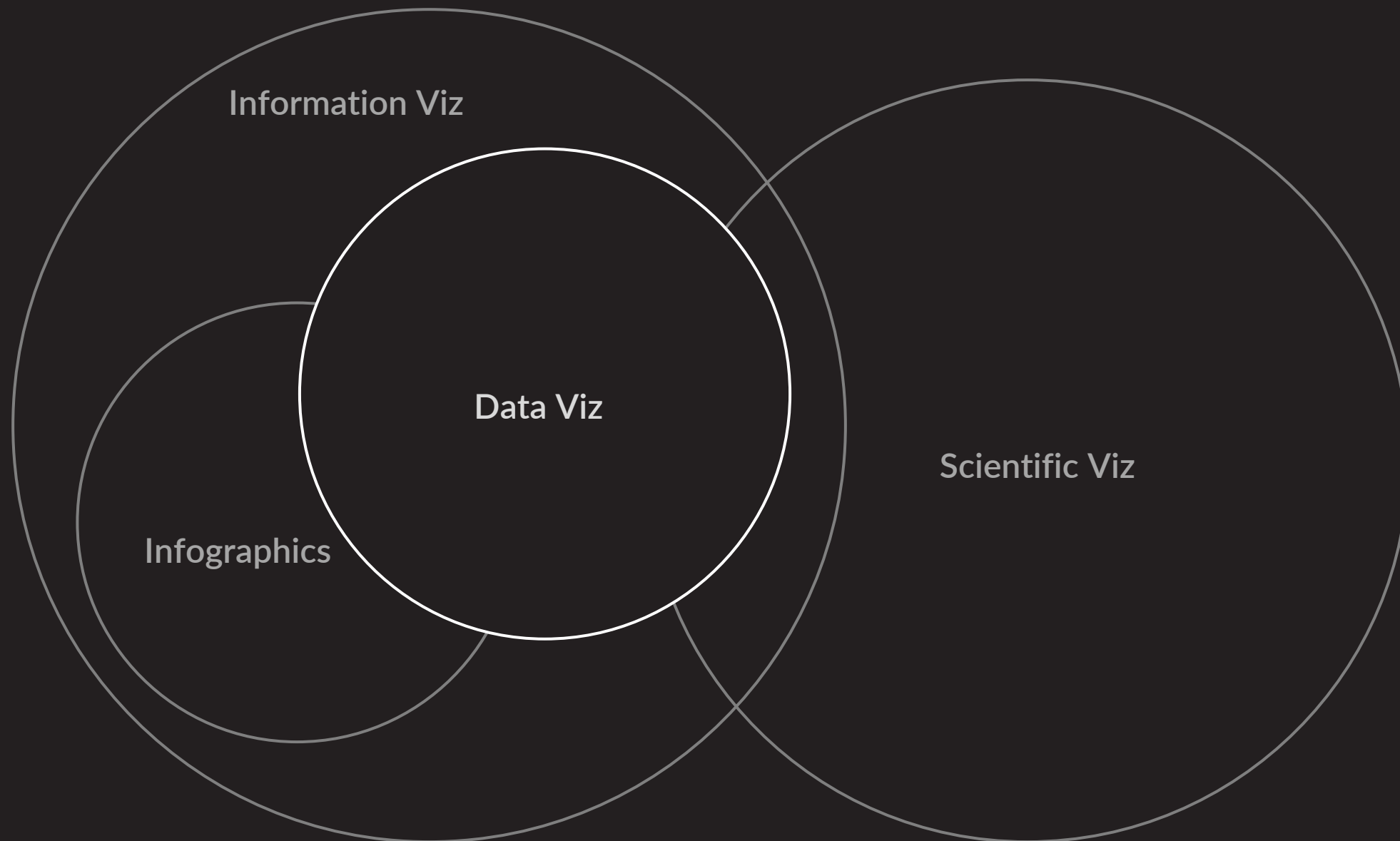
**Data visualization** is a combination of **numerous fields** whose purpose is to communicate information **visually** with a specific **goal** to a specific target **audience**.

Customer  
Company  
Government  
Researcher

...

Video  
Static Figure  
Animated Graph  
Interactive Web

...





mit  
media  
lab



affective  
computing



biomechanics



camera  
culture



changing  
places



civic  
media



design  
fiction



fluid  
interfaces



human  
dynamics



lifelong  
kindergarten



macro  
connections



mediated  
matter



molecular  
machines



object-based  
media



opera of  
the future



personal  
robots



playful  
systems



responsive  
environments



social  
computing



social  
machines



speech +  
mobility



synthetic  
neurobiology

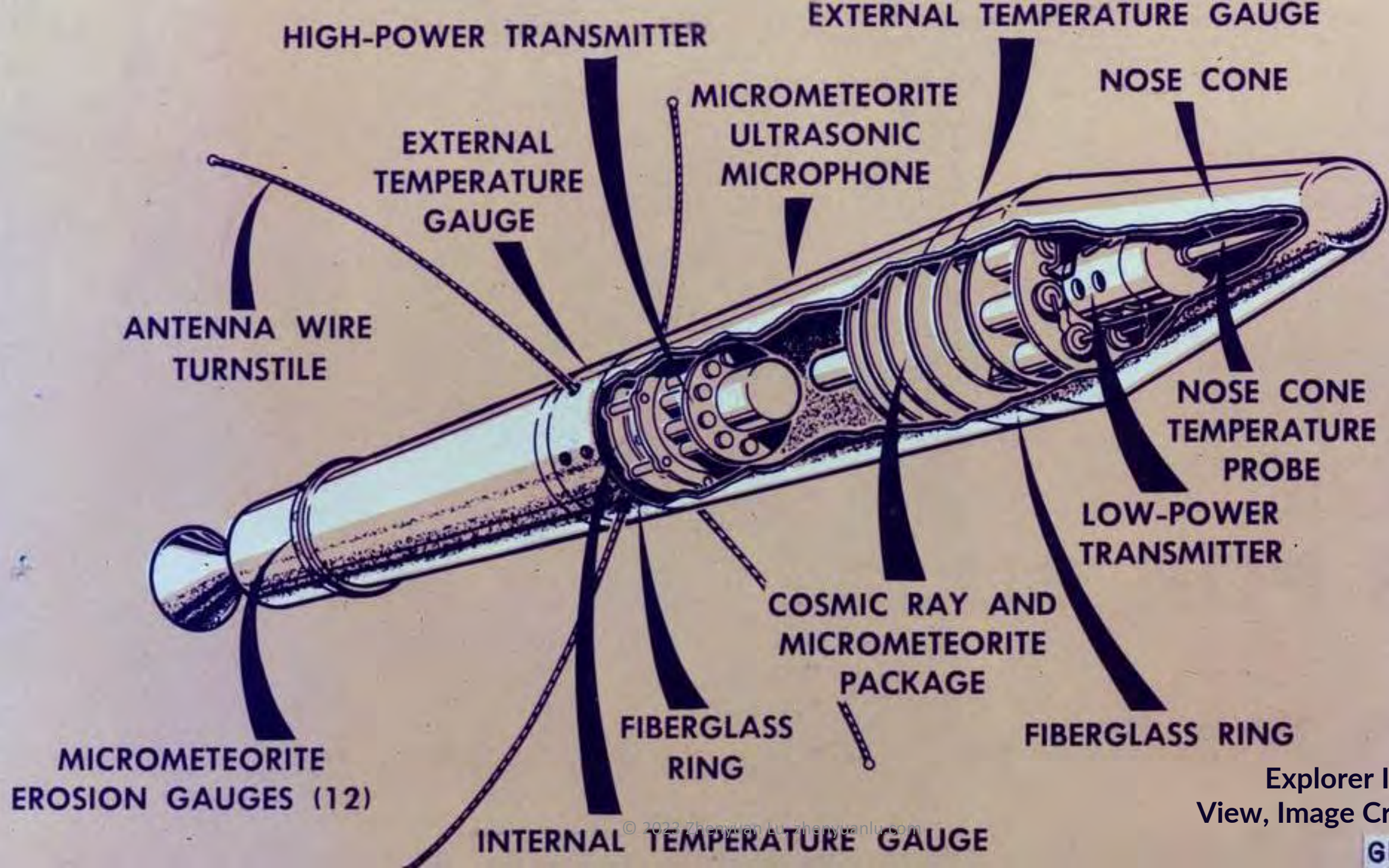


tangible  
media



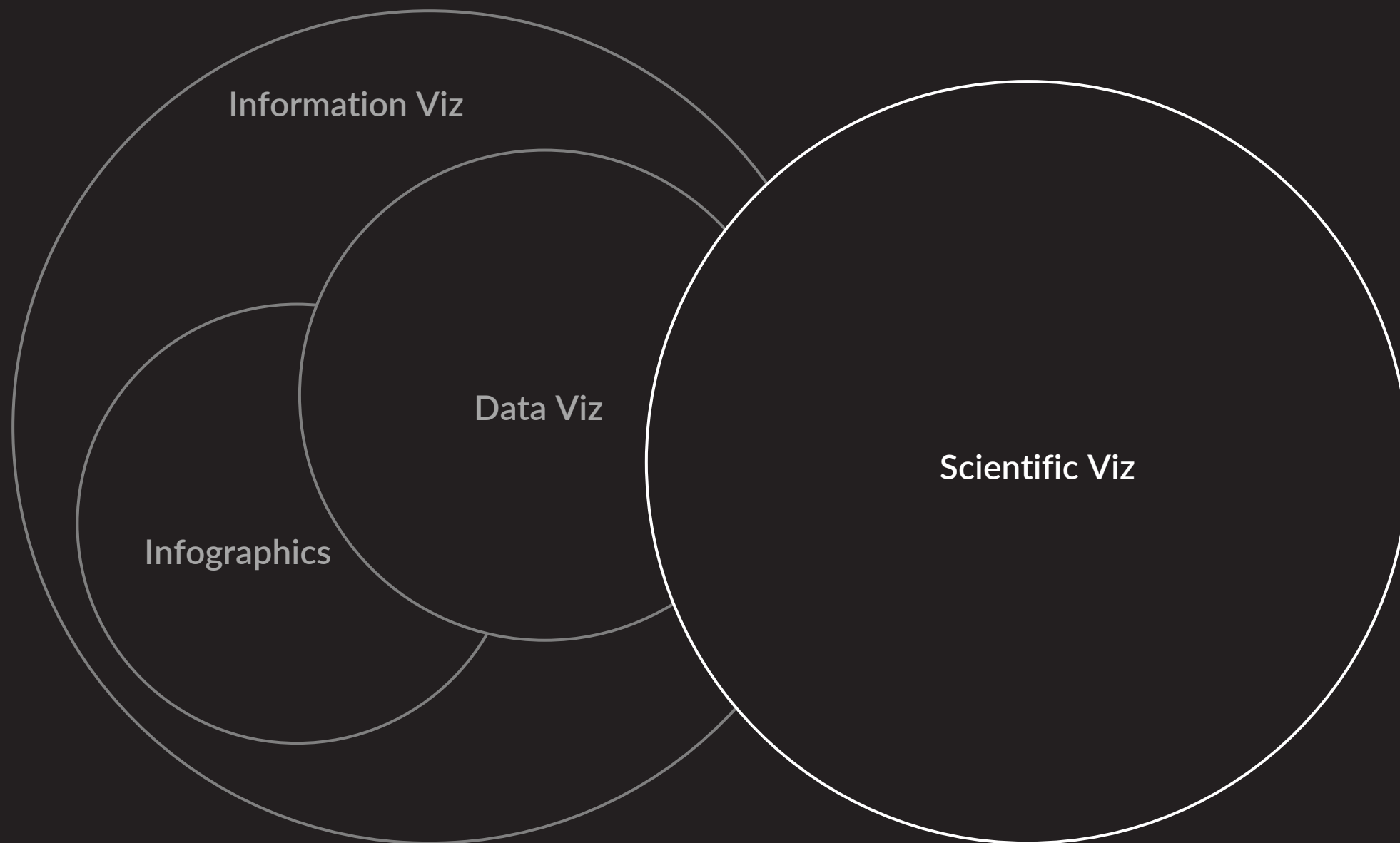
viral  
communications

Michael Bierut of Pentagram designed MIT Media Lab's new logo



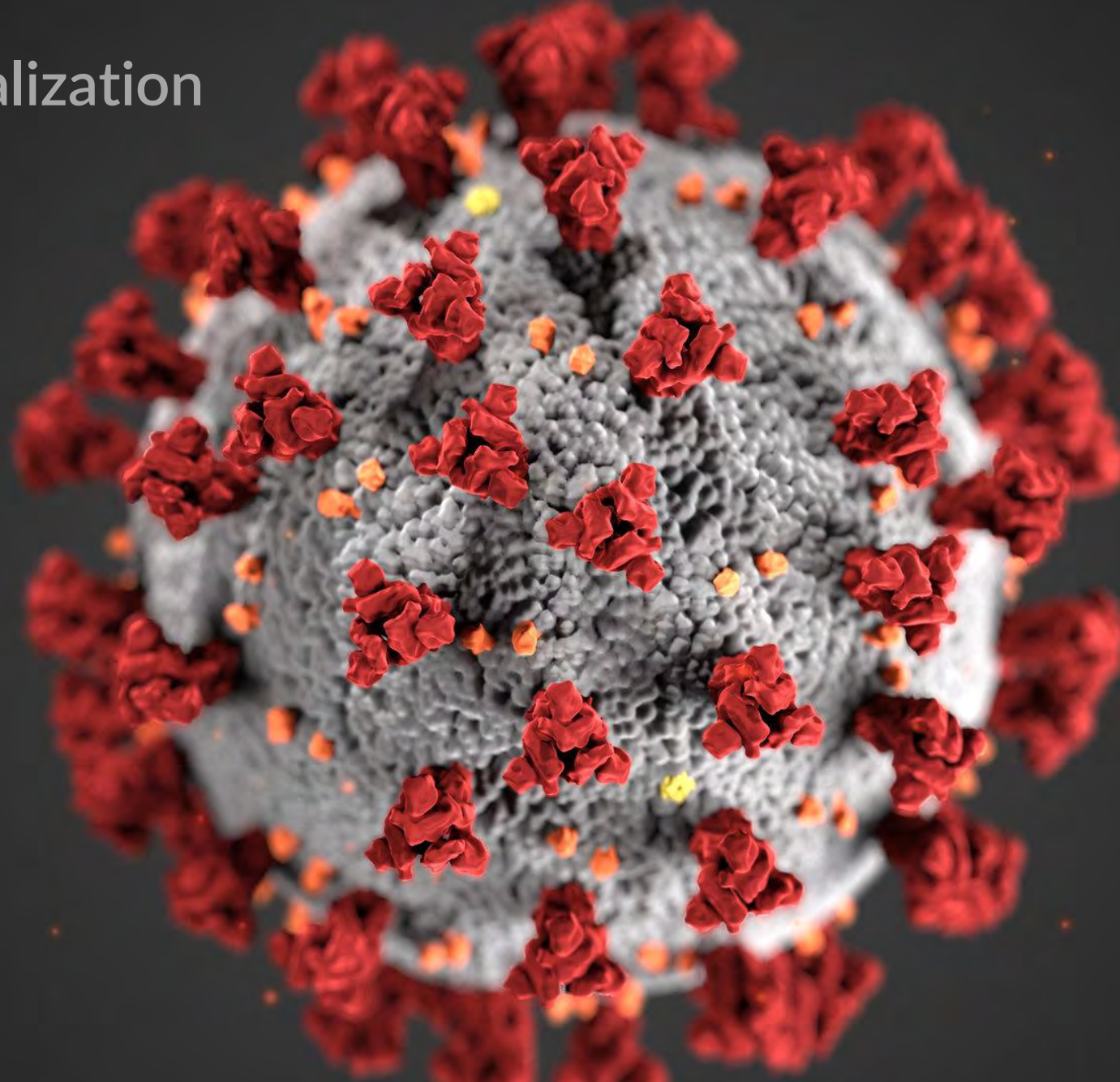
Explorer I - Exploded View, Image Credit: NASA

GE 142-59





# Scientific Visualization

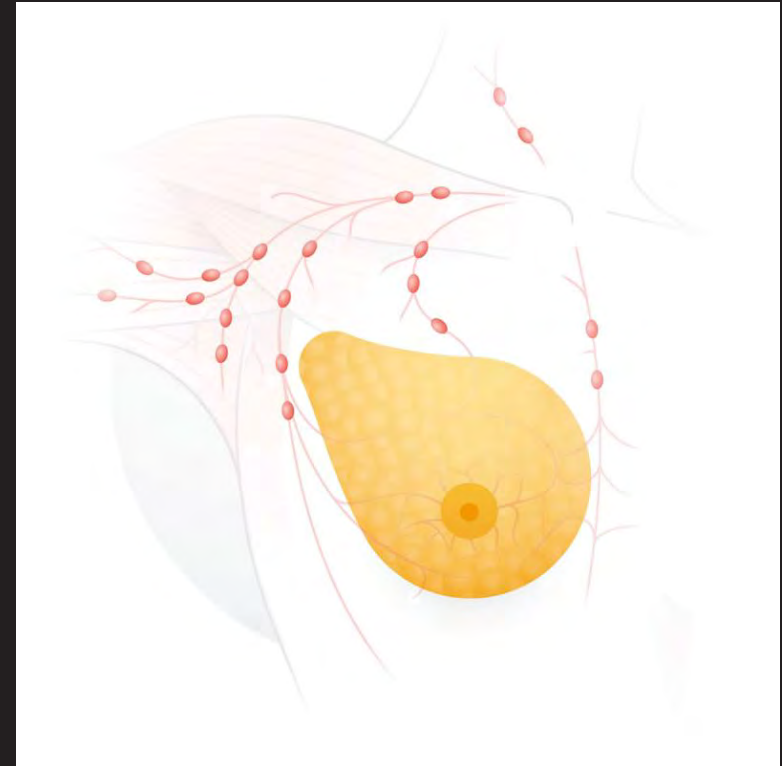




# Scientific Visualization

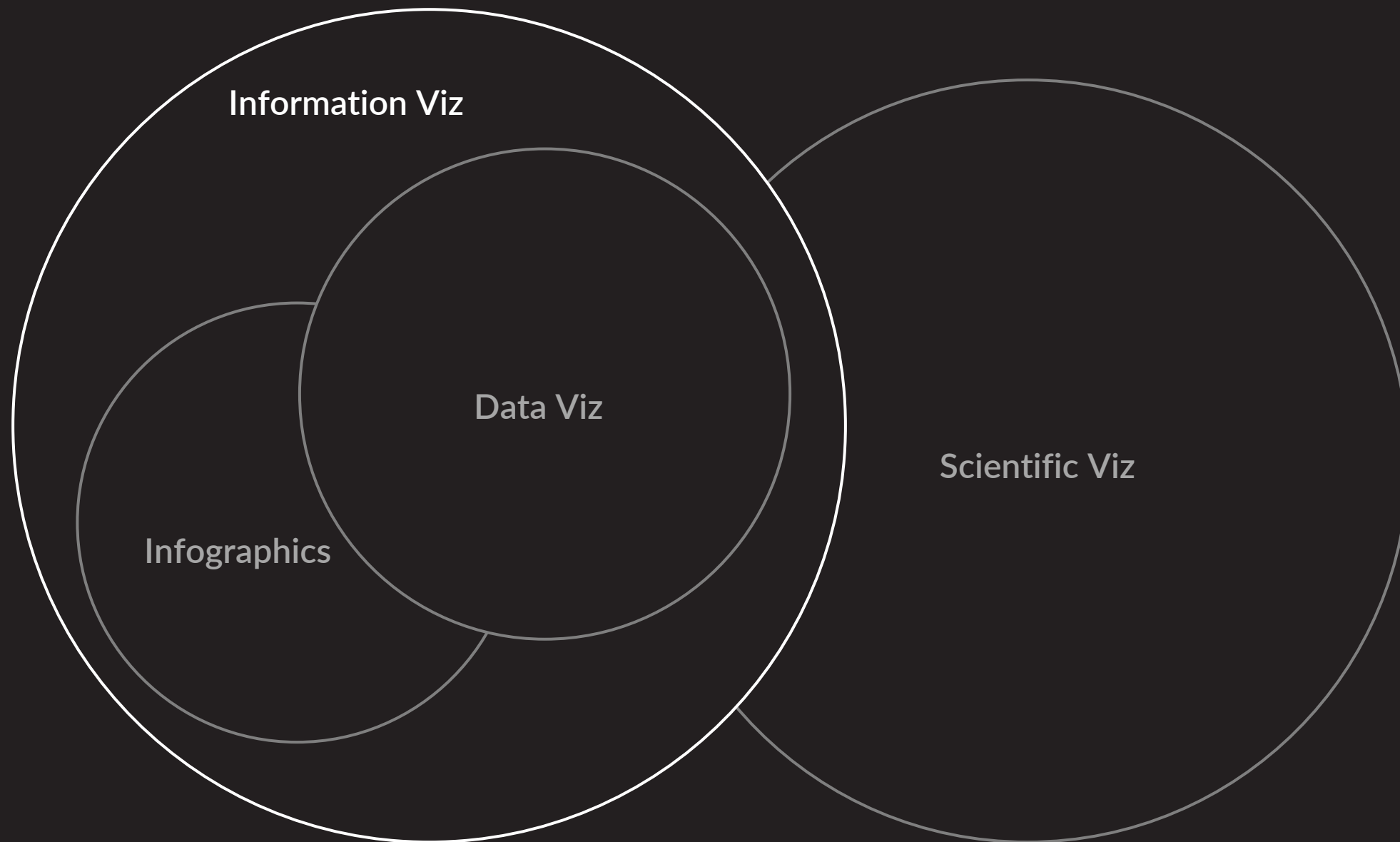


Initial scan with AI detection



Ardila, D., et al., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography [\[link\]](#), *Nature Medicine*, Vol 25(6), pp. 954--961. Springer Science and Business Media {LLC}. DOI: [10.1038/s41591-019-0447-x](https://doi.org/10.1038/s41591-019-0447-x)

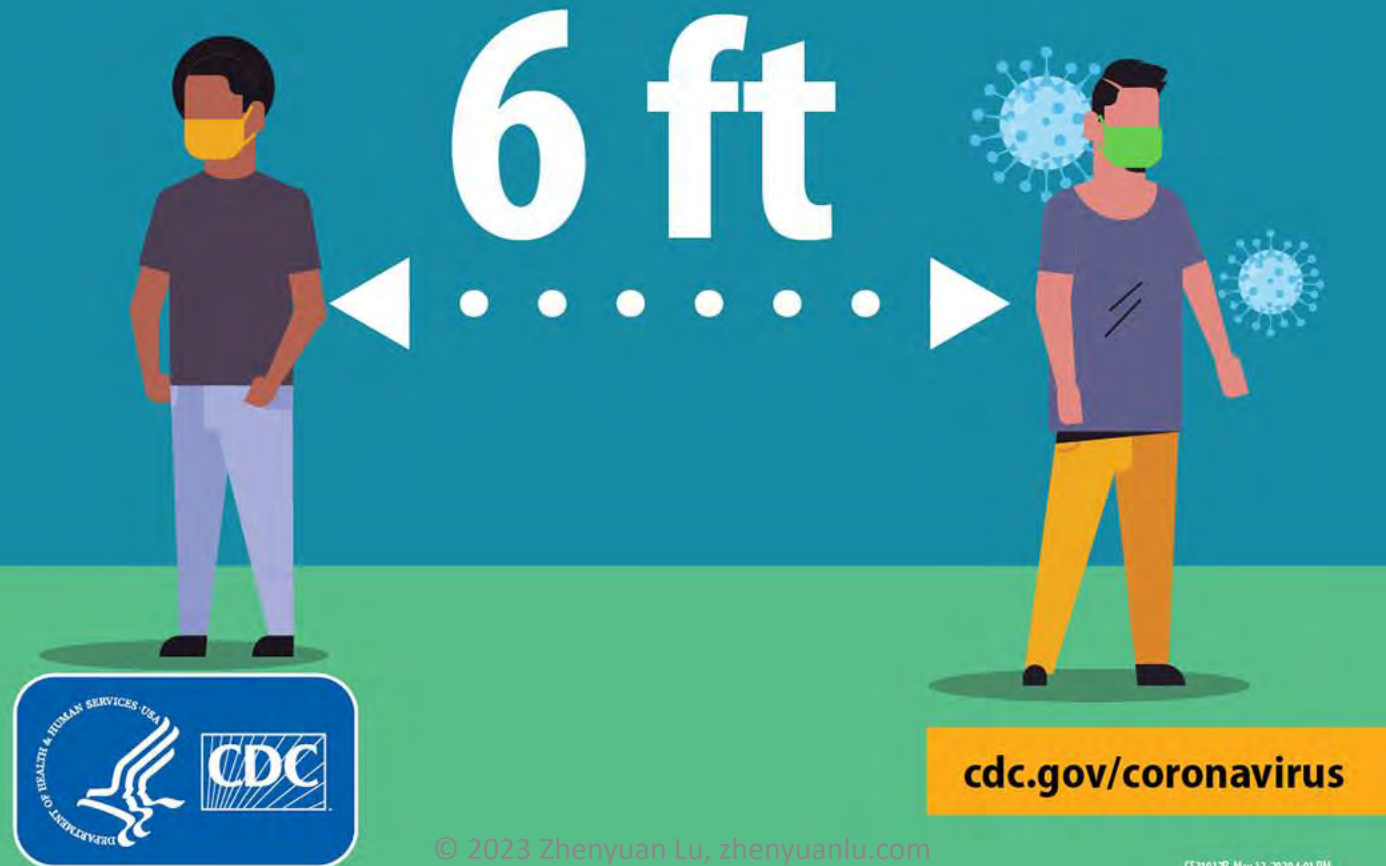
McKinney, S.M., et al., 2020. International evaluation of an {AI} system for breast cancer screening [\[link\]](#), *Nature*, Vol 577(7788), pp. 89--94. Springer Science and Business Media {LLC}. DOI: [10.1038/s41586-019-1799-6](https://doi.org/10.1038/s41586-019-1799-6)



# STOP THE SPREAD OF GERMS

Help prevent the spread of respiratory diseases like COVID-19.

Stay at least 6 feet (about 2 arms' length)  
from other people.

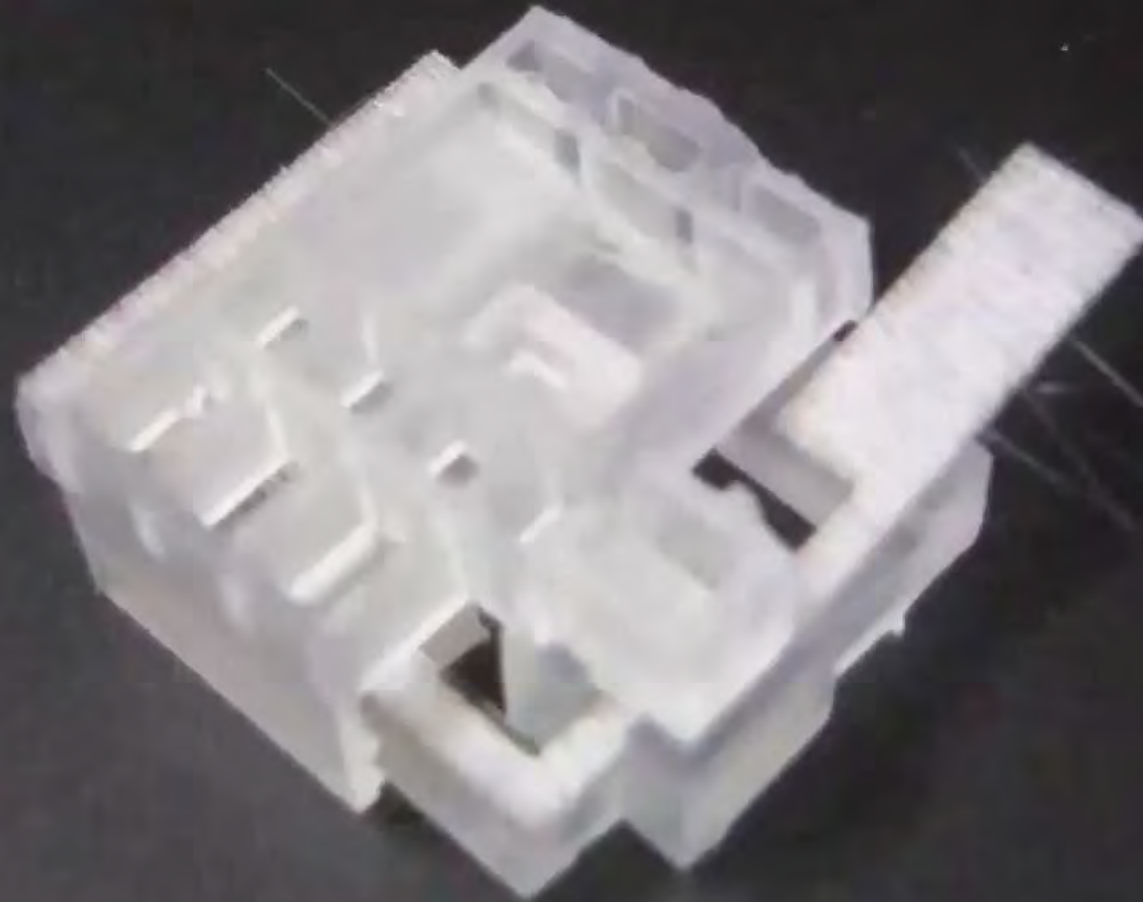




Serpentine  
structure

10xSpeed  
mm:ss

00;00;00;00



- [1] Zhenyuan Lu , et. al (2018), Point-of-care Diagnostic Cartridge Having A Programmable Fluidic Wicking Network, US Patent
- [2] Kshitij Ranjan, et. al (2017), Development of a paper-based diagnostic platform for the detection of diarrhea causing pathogens, BMES
- [3] Zhenyuan Lu , et. al (2016), Paper-based Device For Gastroenteritis Detection Integrated With Sample Preparation Cartridge, BMES



# Infographic

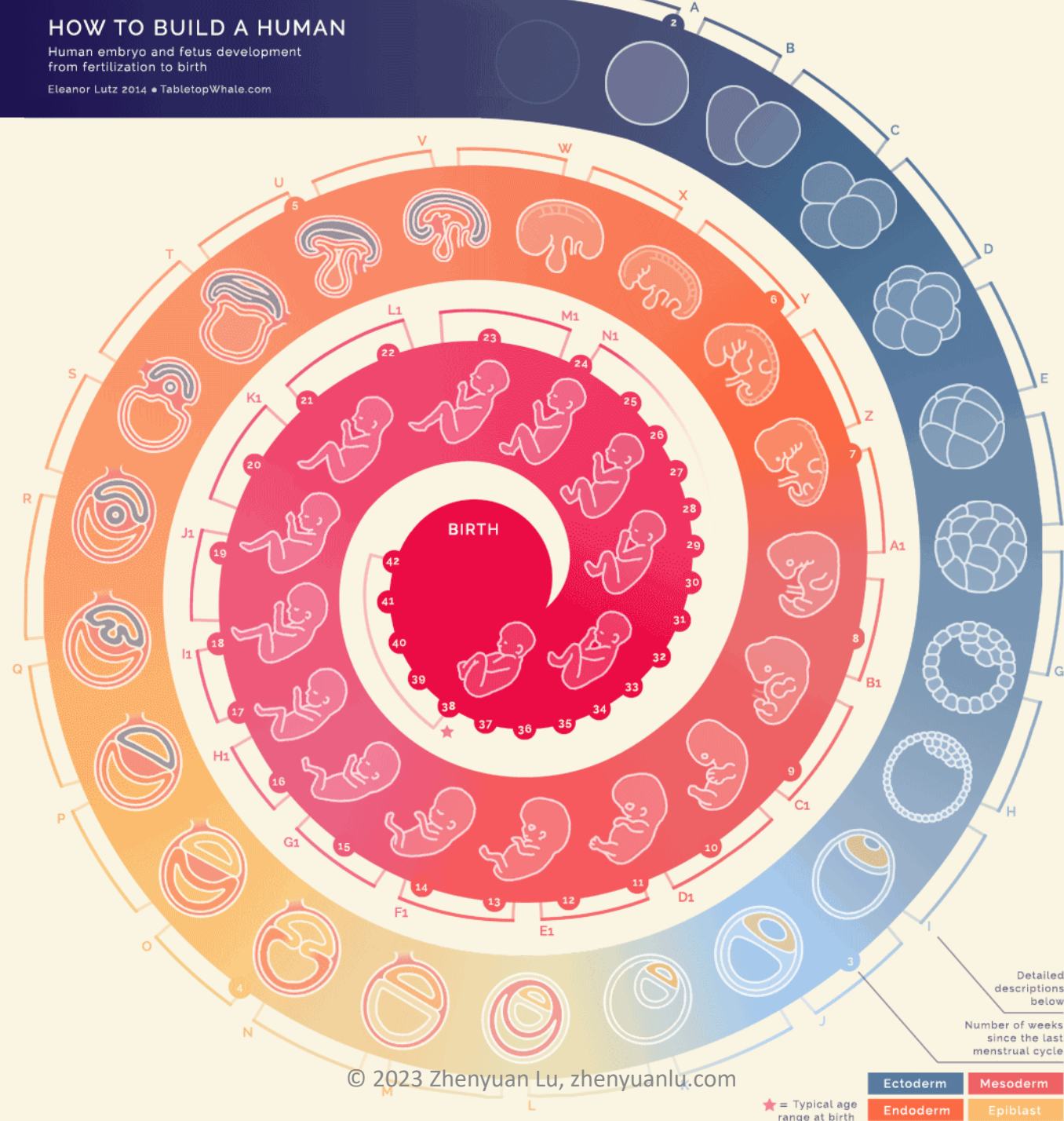


# Infographic

## HOW TO BUILD A HUMAN

Human embryo and fetus development  
from fertilization to birth

Eleanor Lutz 2014 • TabletopWhale.com



© 2023 Zhenyuan Lu, zhenyuanlu.com

Detailed descriptions below  
Number of weeks since the last menstrual cycle

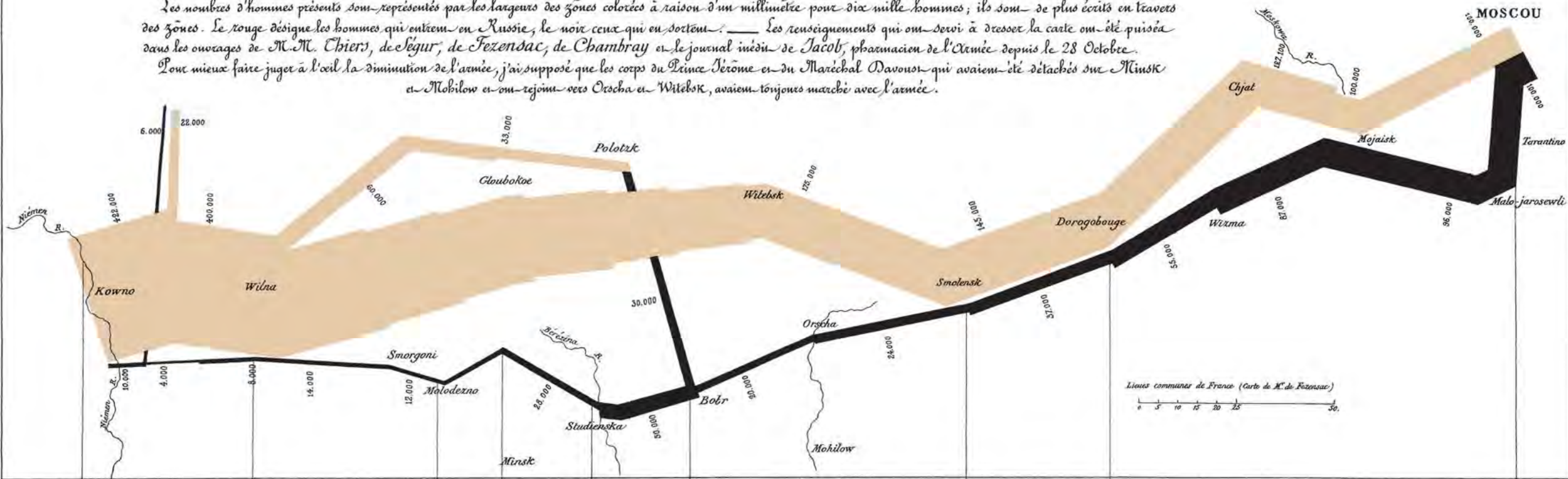
Author: Eleanor Lutz, 2014  
Source: Gilbert, Scott. Developmental Biology, 9th Edition. Sunderland, MA: Sinauer Associates Inc., 2010.



## Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

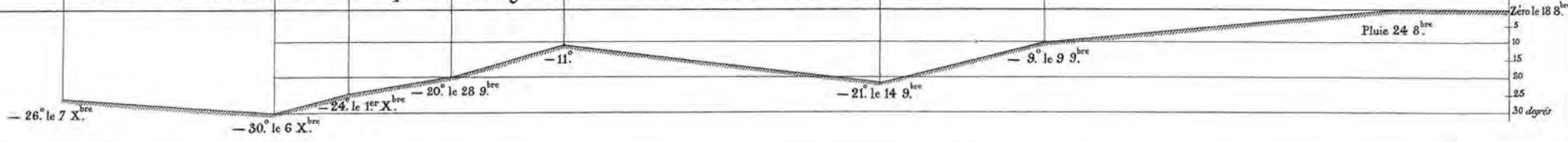
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Fozensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.



### TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen gelé.



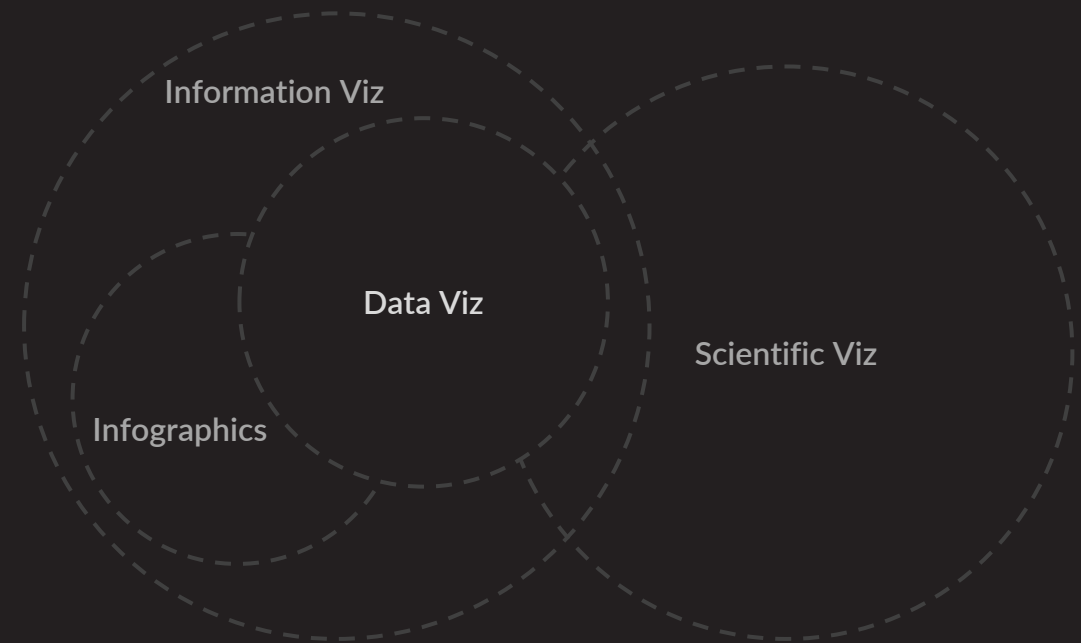
Autog. par Regnier, 8. Pas. 5<sup>66</sup> Marie 5<sup>66</sup> à Paris.

Imp. Lith. Regnier et Douard.





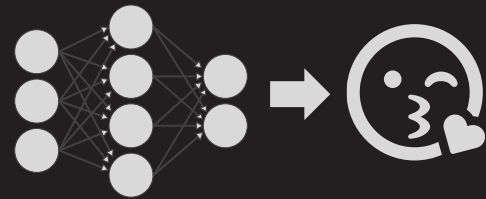
Don't be limited by these. Consider all of these circles to be tools that will assist you in achieving your **objectives**.



# Why Data Visualization

Data is everywhere.





Data is **everywhere**.

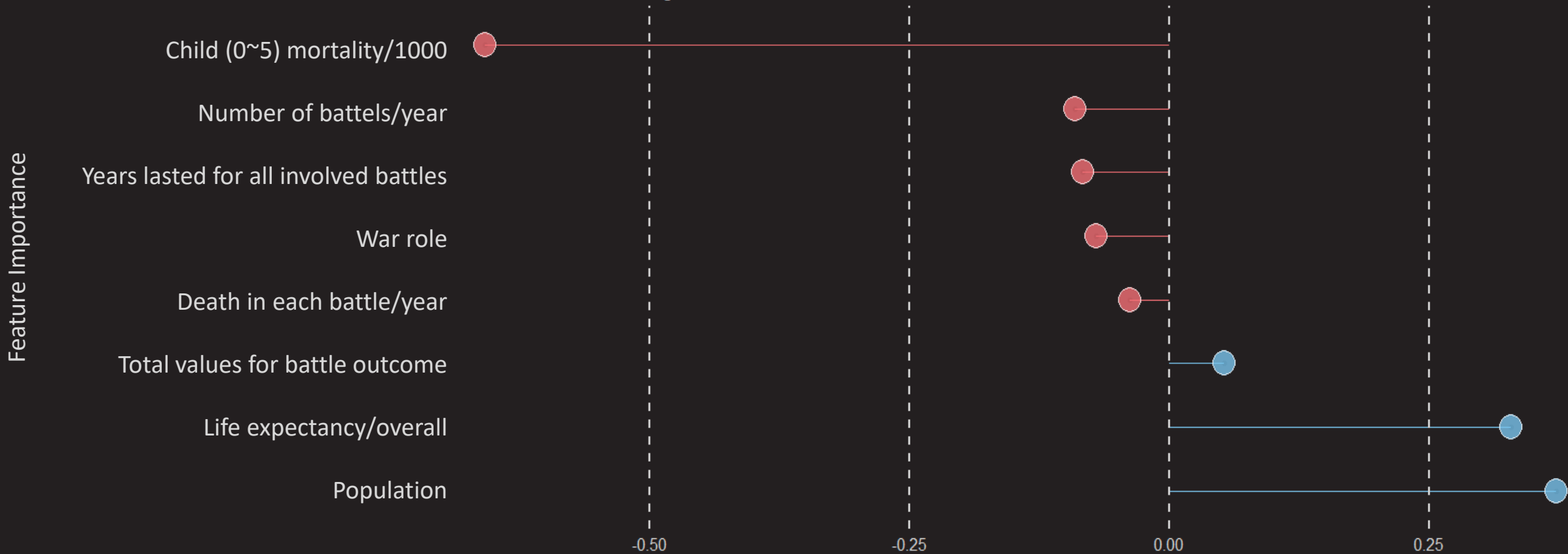


Unstructured information is **hard** to understand.

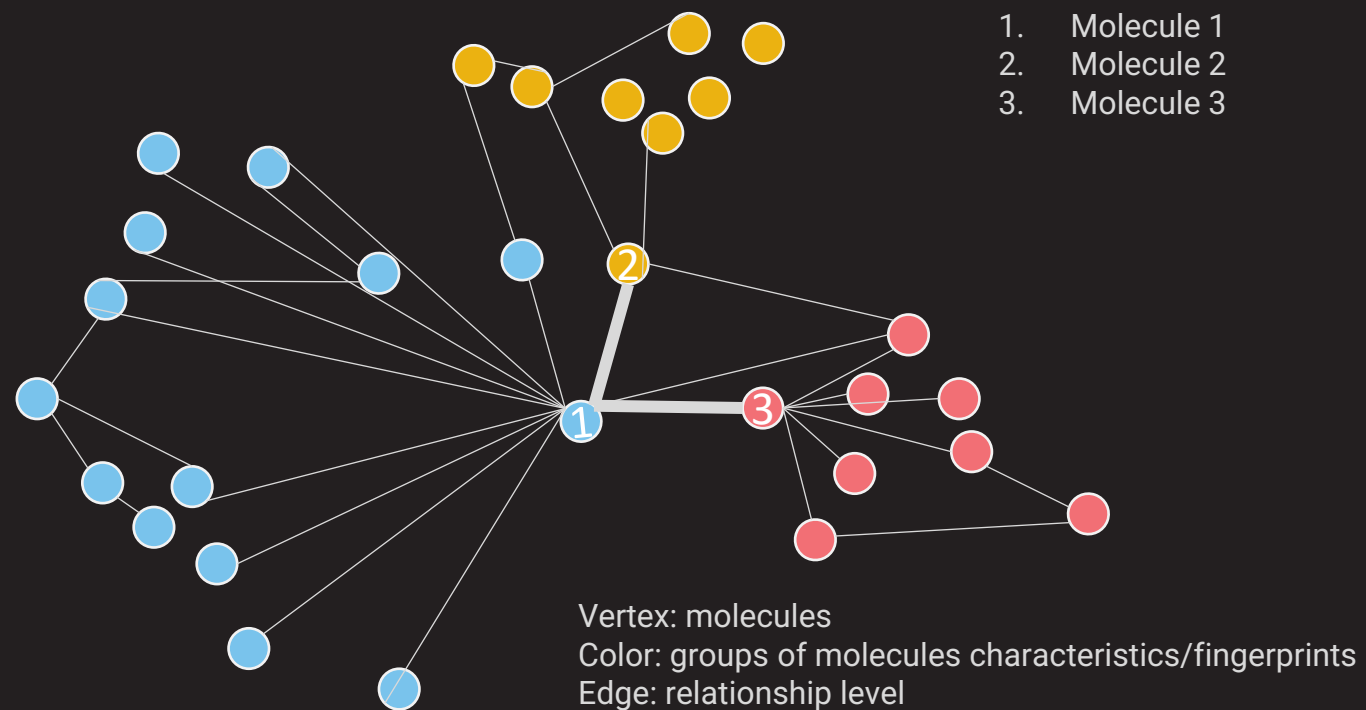
*and sometimes boring...*

# Comparison

Centralized Degree Correlation Analysis  
Positive Correlations vs. Negative Correlations

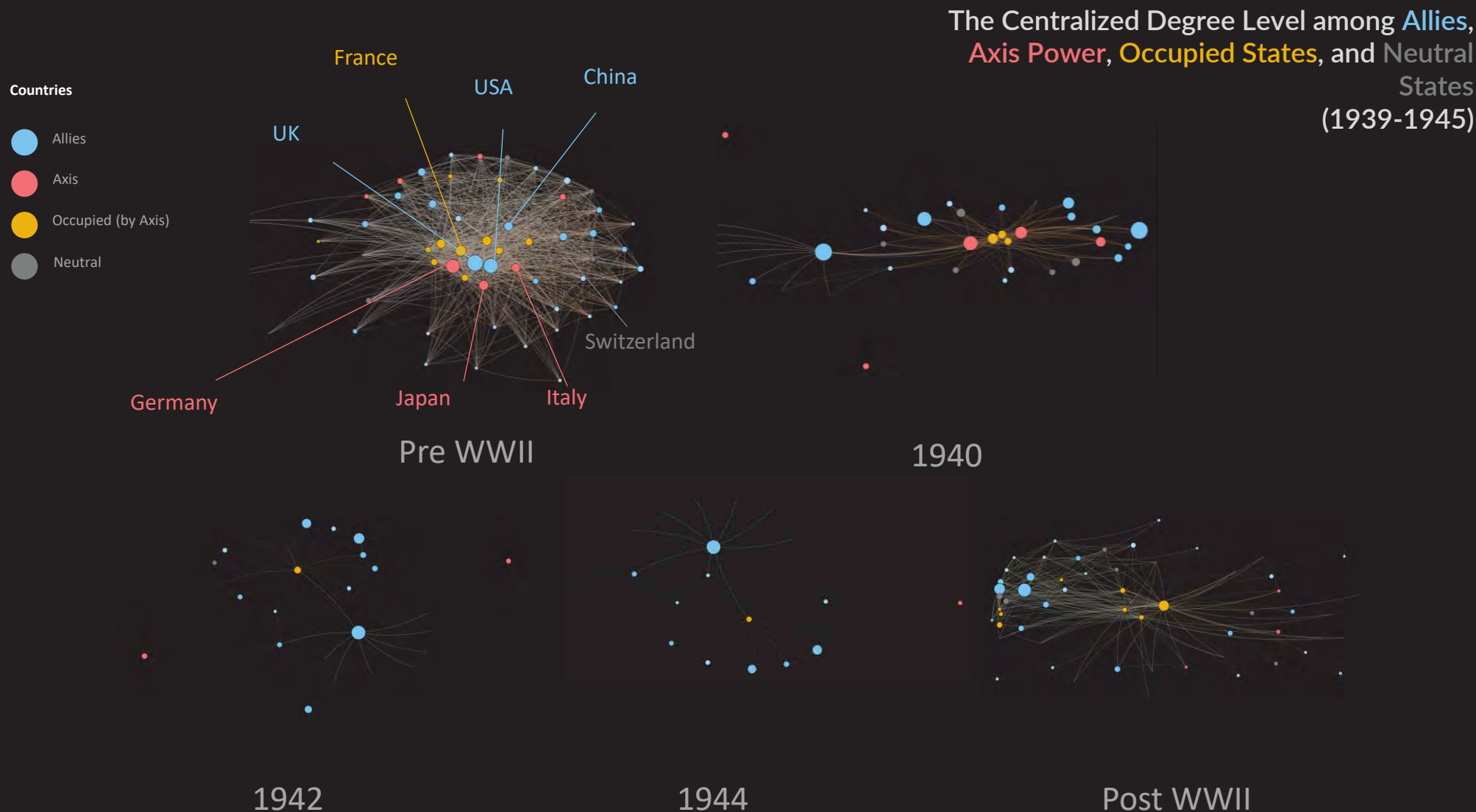


# Understand Relationship



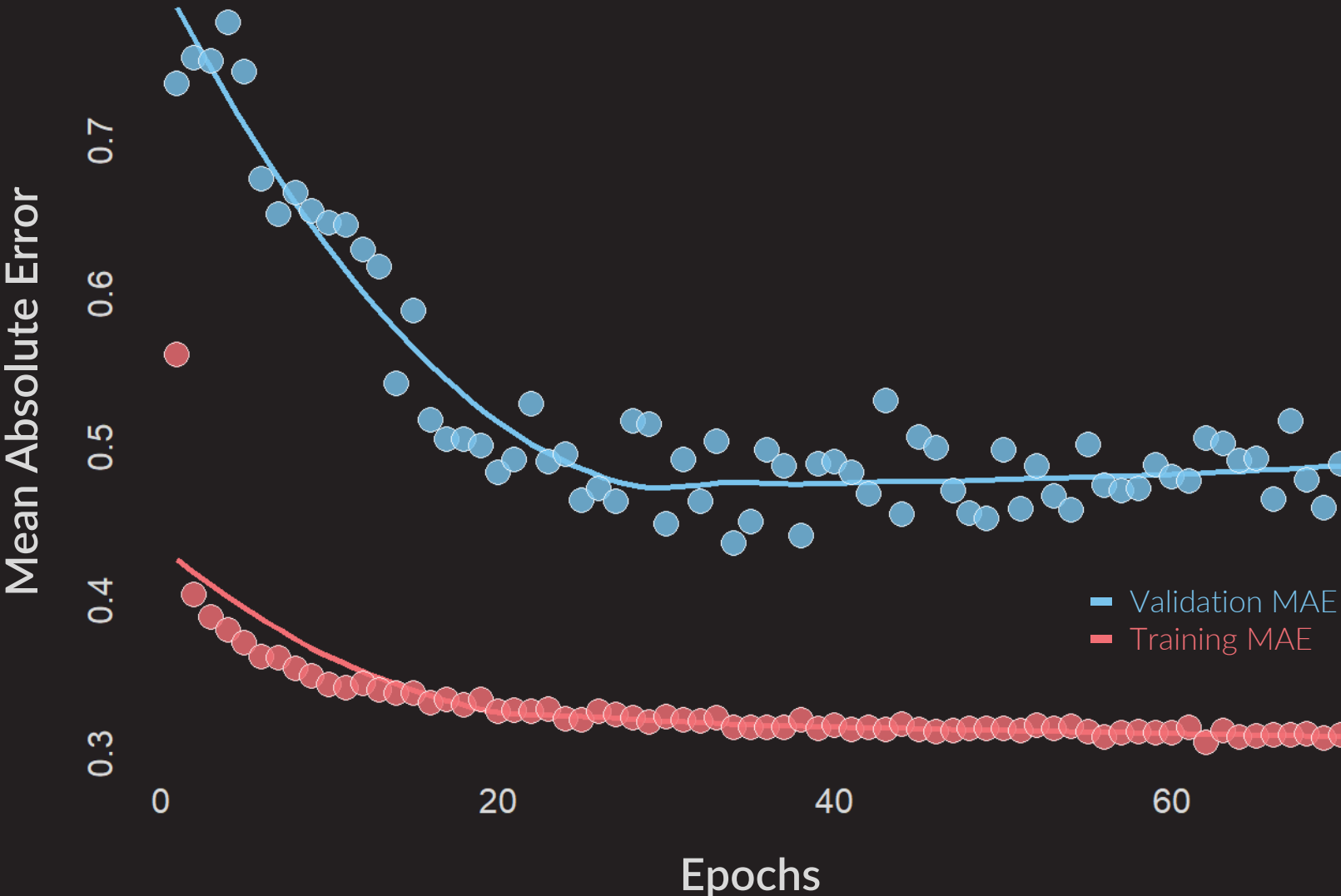


# Observation of Our World

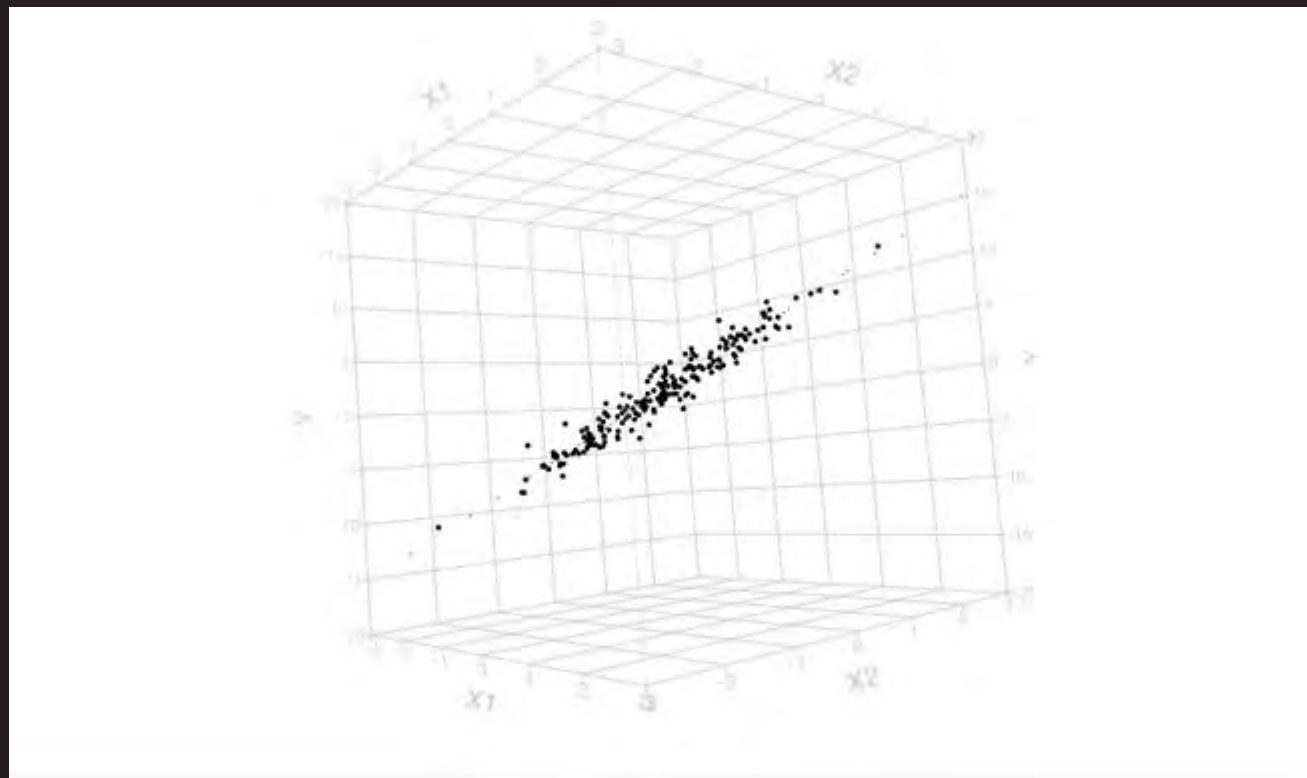


# Investigation of Model Performance

Results



# Detection on Patterns

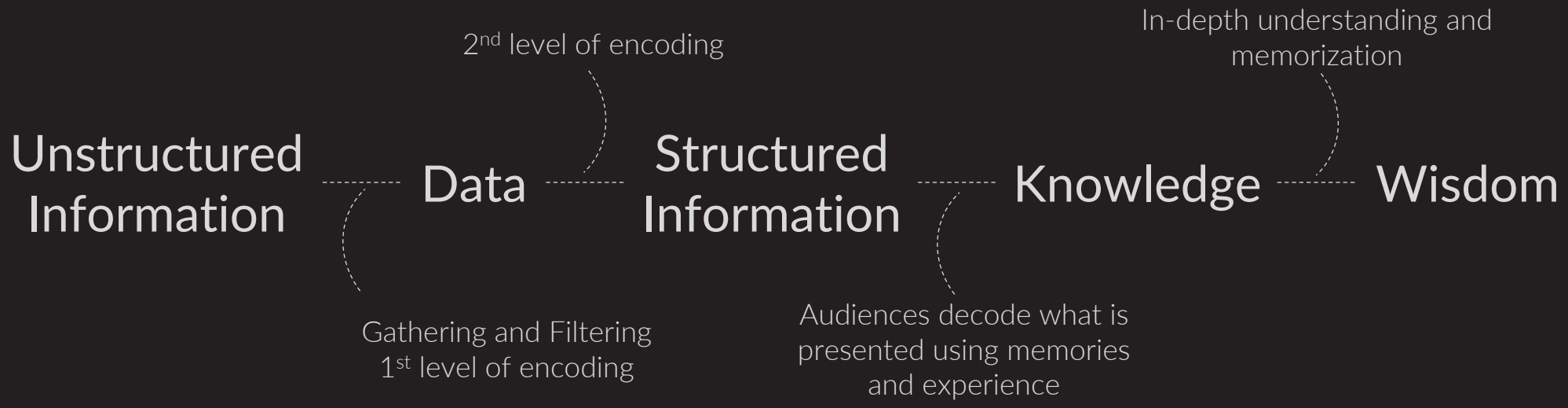
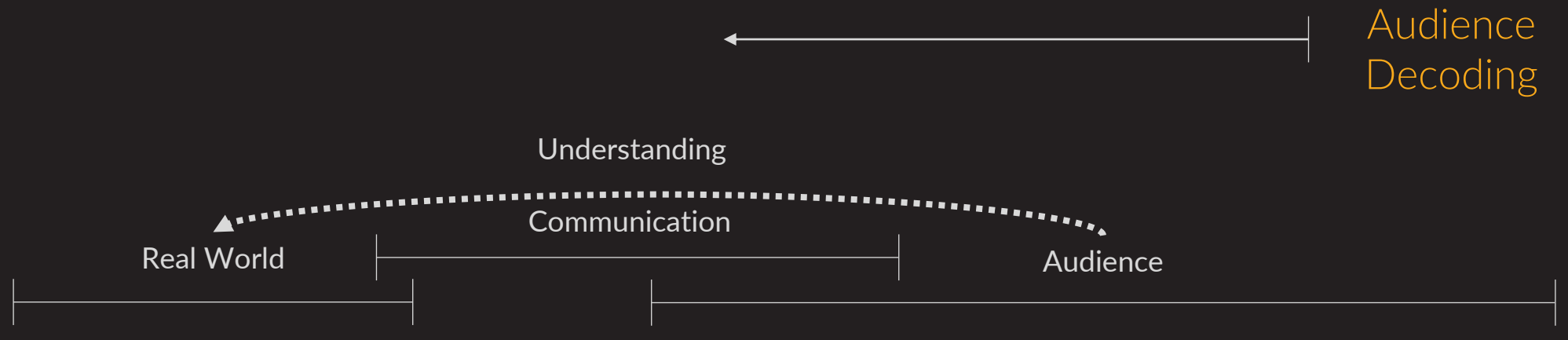


# How to Build Data Visualization

Know your audience.

Know your data, respect your audience.

David Alayón



## Unstructured Information



Image Credit: Newton's Apple

## Structured Information

$$F = G \frac{m_1 m_2}{r^2}$$



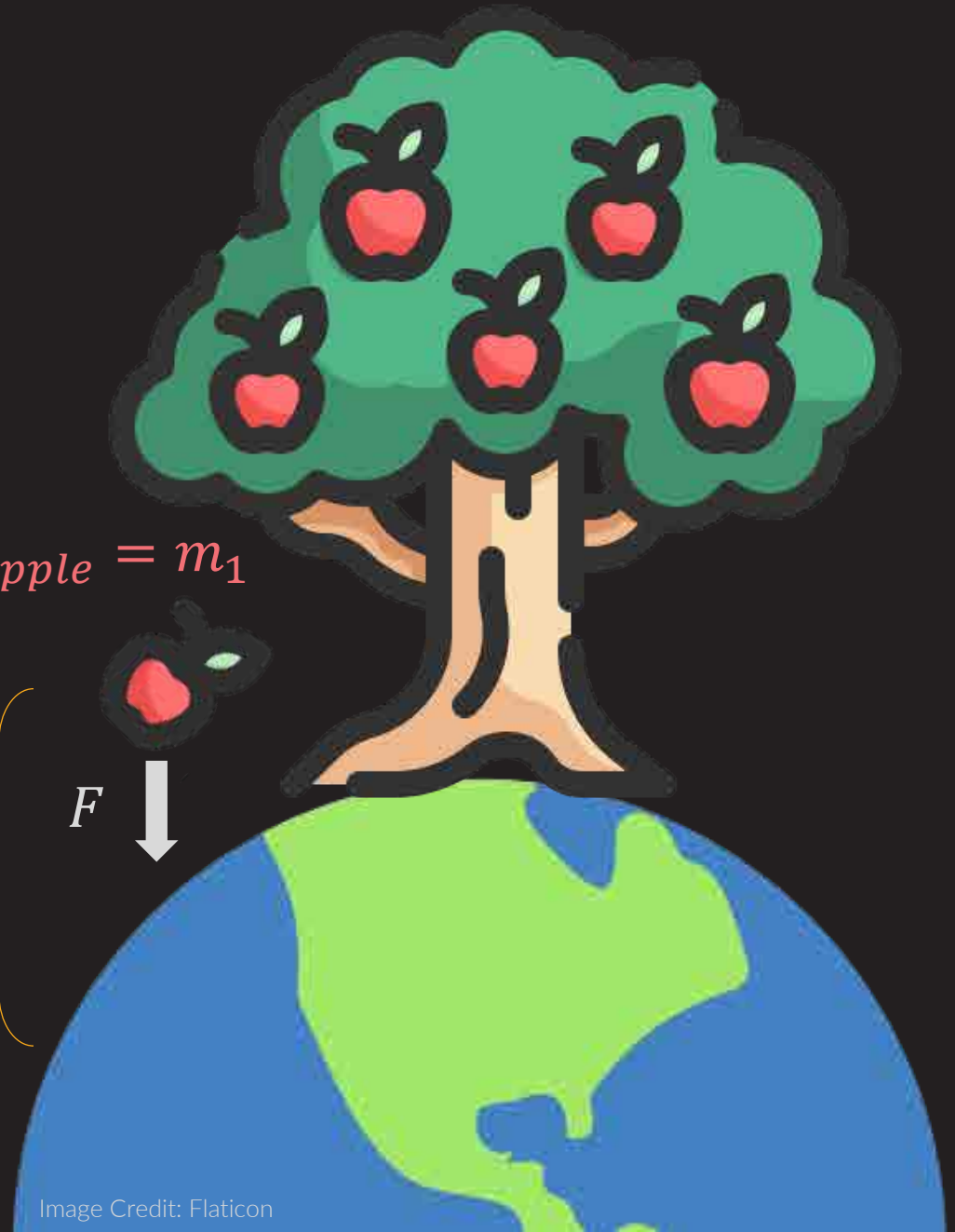
# Newton's Law of Gravitation

$$F = G \frac{m_1 m_2}{r^2}$$

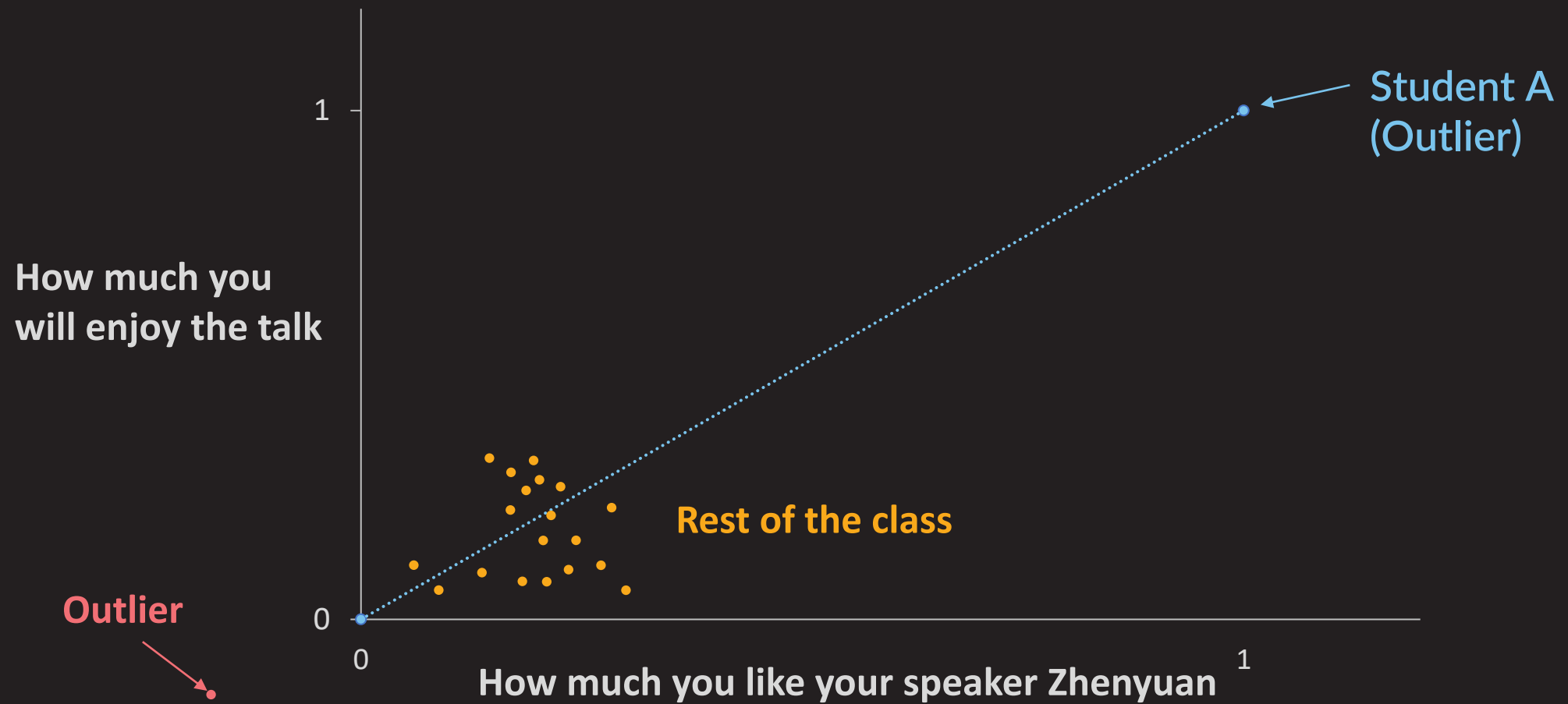
$Mass_{apple} = m_1$

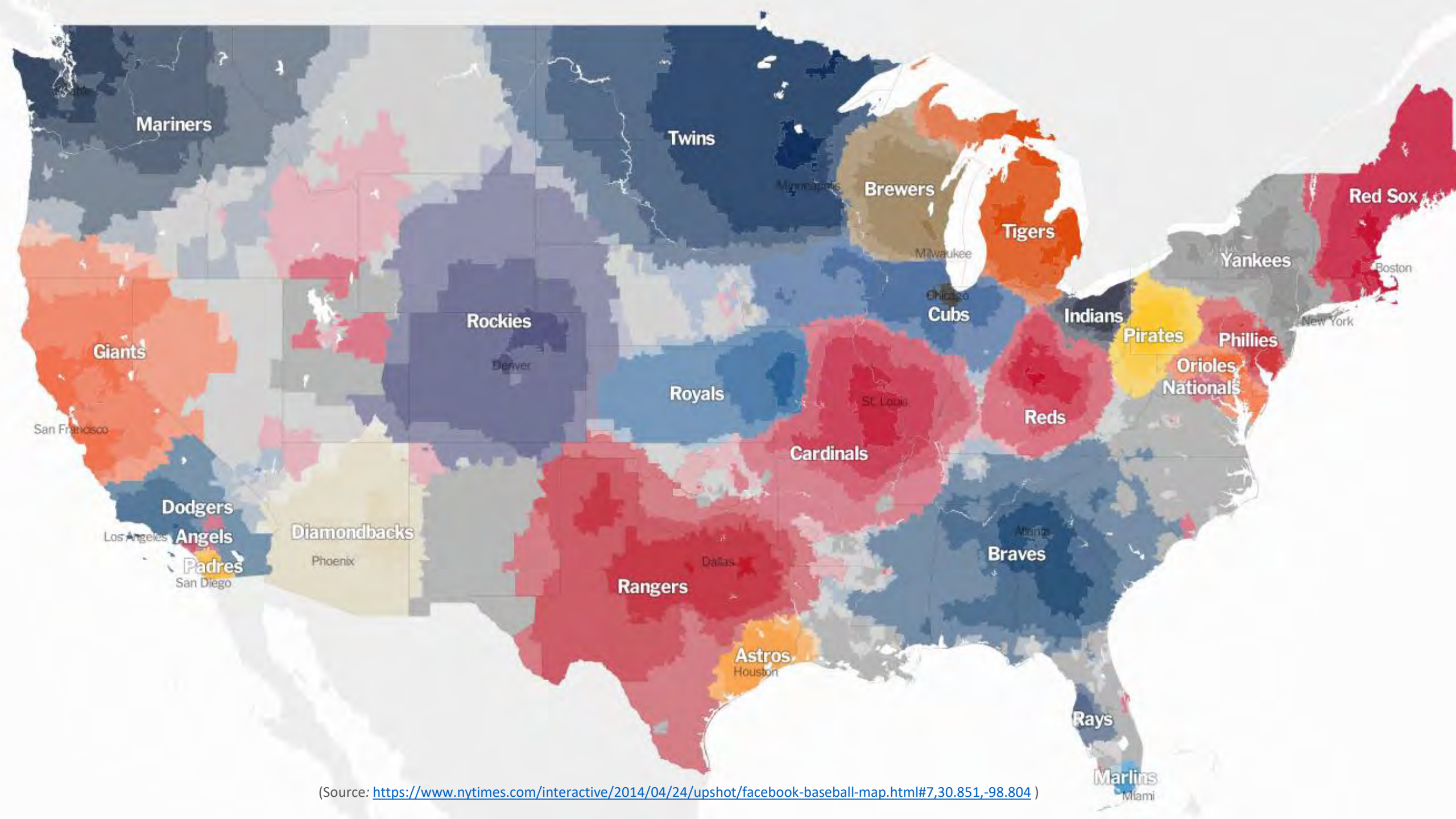
$Distance = r^2$

$Mass_{earth} = m_2$



# Could you decode this chart?

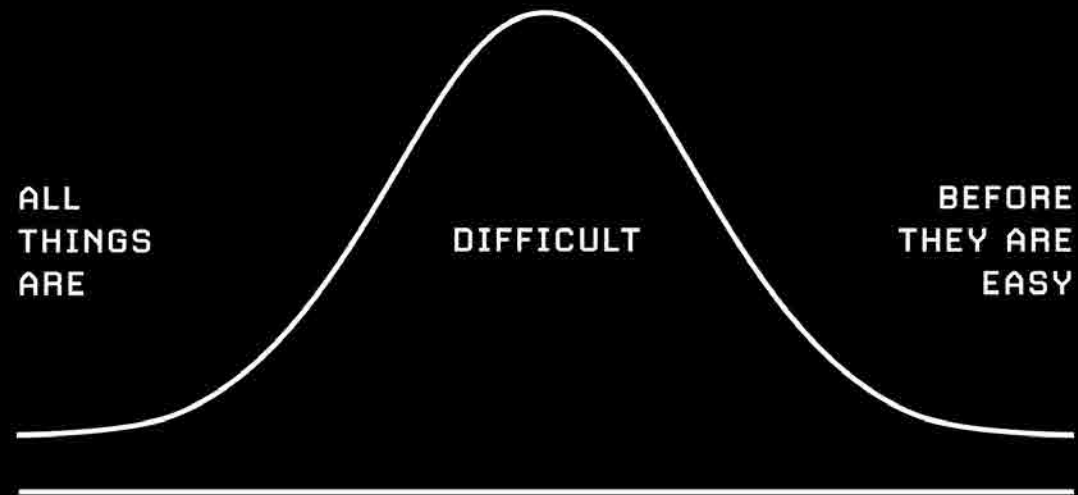




(Source: <https://www.nytimes.com/interactive/2014/04/24/upshot/facebook-baseball-map.html#7,30.851,-98.804> )



Image Credit: Ryan MacEachern



*"All things are difficult before they are easy." – Chinese Proverb*

万事开头难



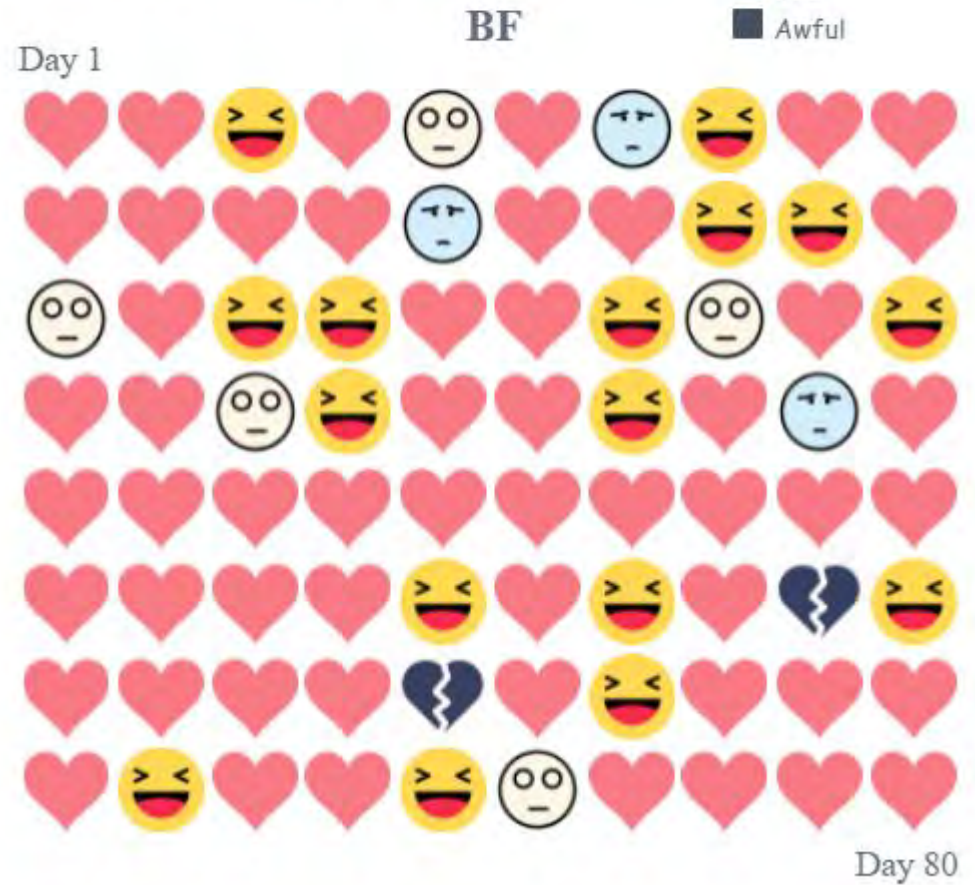
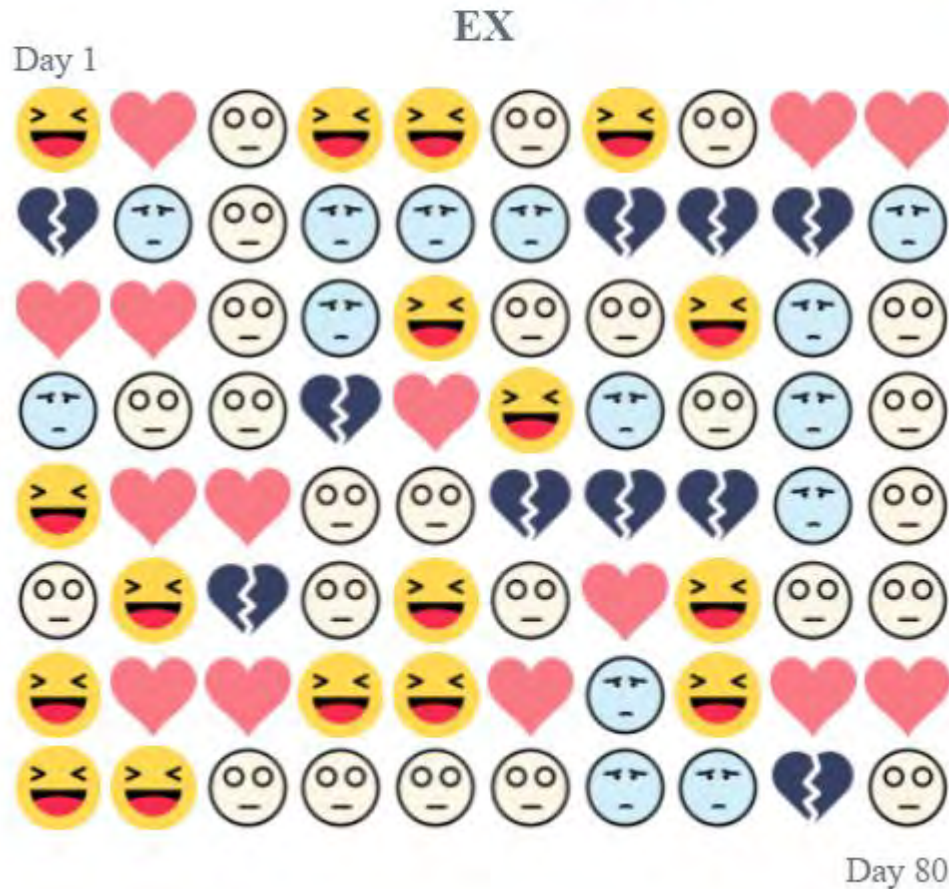
# How do I tell if I'm dating the right guy?

5  
moods

80  
days

1,064  
% of improvement

- Moods Recorded
- Amazing
  - Happy
  - Okay
  - Sad
  - Awful



### Crown favorites : trying to figure out who will be the next king?

hover for more details or click to access the family tree



### Survival probability study :

Jon and Arya are the one most likely to survive.



### What is on George RR Martin's head?

Based on the Nb of lines by characters and season during the show, Tyrion is a very popular character and the one who spoke the most.



### Based on military strenght :

Daenerys seems to have a serious asset...



### Special asset



### What if the King was elected by twitter?

Nb of tweets in bars and retweets in color; Arya is at the center of many discussion, but Jon followers push it harder...



### Ranking in term of lines told by characters and season

Jon became more popular through the seasons, and at the end of season 6, he achieved the first rank.



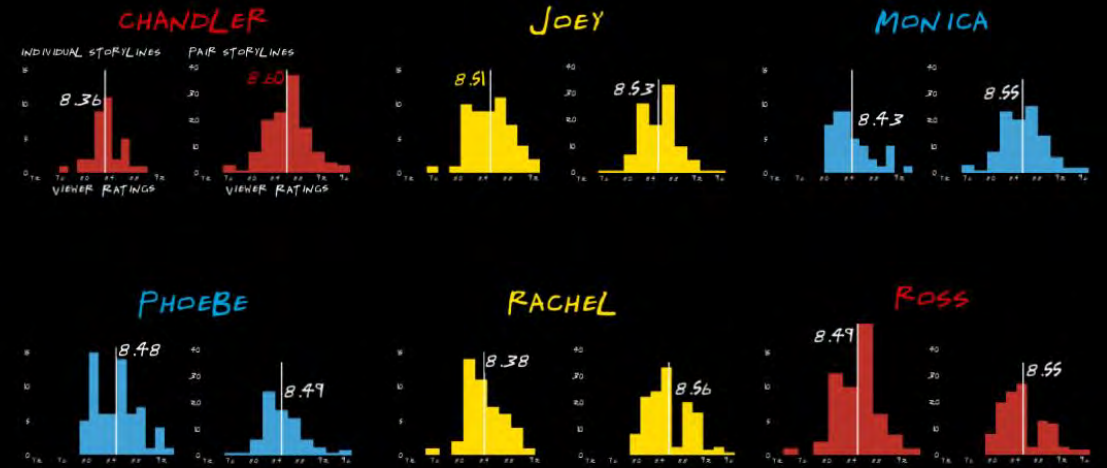
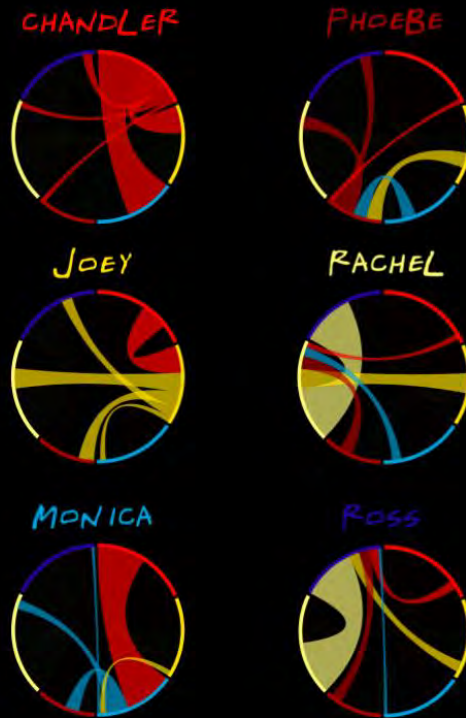
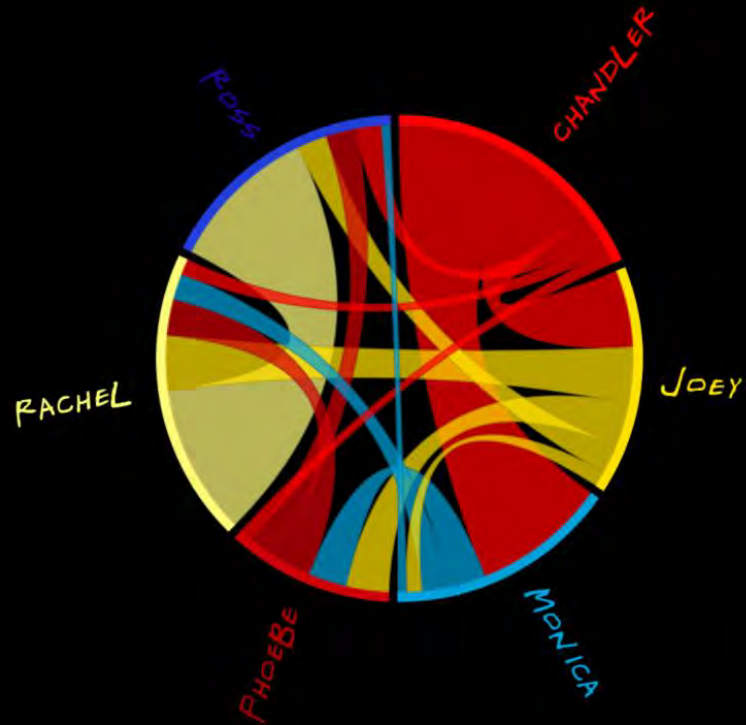
hover to see the credits

Based on the survival probability, the army strenghts and the increased popularity of his character on the show, the next Iron king will be : Jon Snow. Because let face it, the Night king can't be the next Iron King.... and he is one of my favourite.





# THE ONE WITH ALL THE RELATIONSHIPS





# One Simple Rule *For most of the cases...*

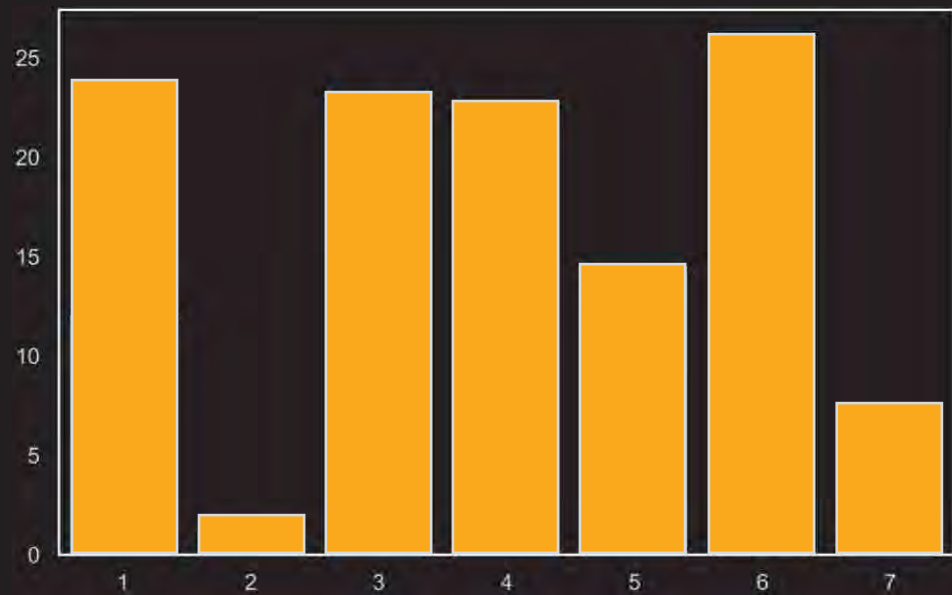
The higher the better

$$\text{Data – ink ratio} = \frac{\text{Ink that encodes data}}{\text{ink used to print everthing else}}$$

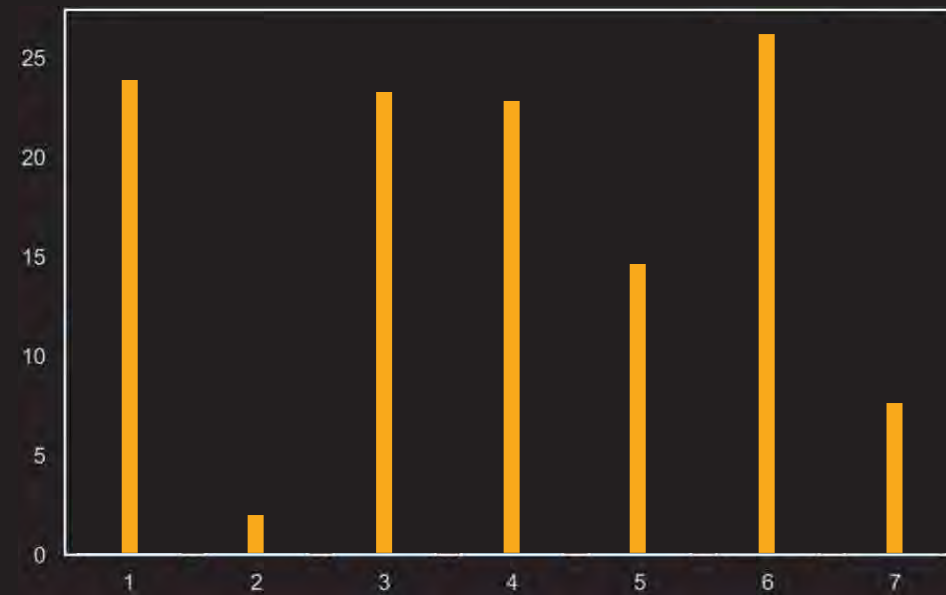
If the dominator is high  
means you focus too  
much on unimportant  
part

Alberto Cairo

## Lower data-ink ratio



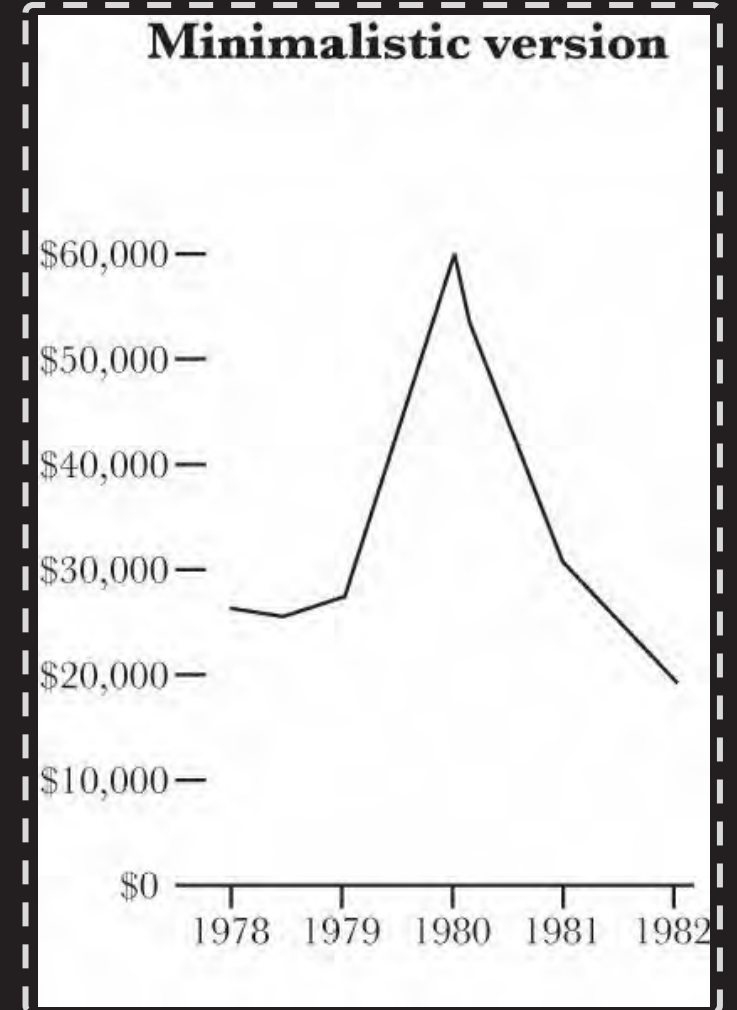
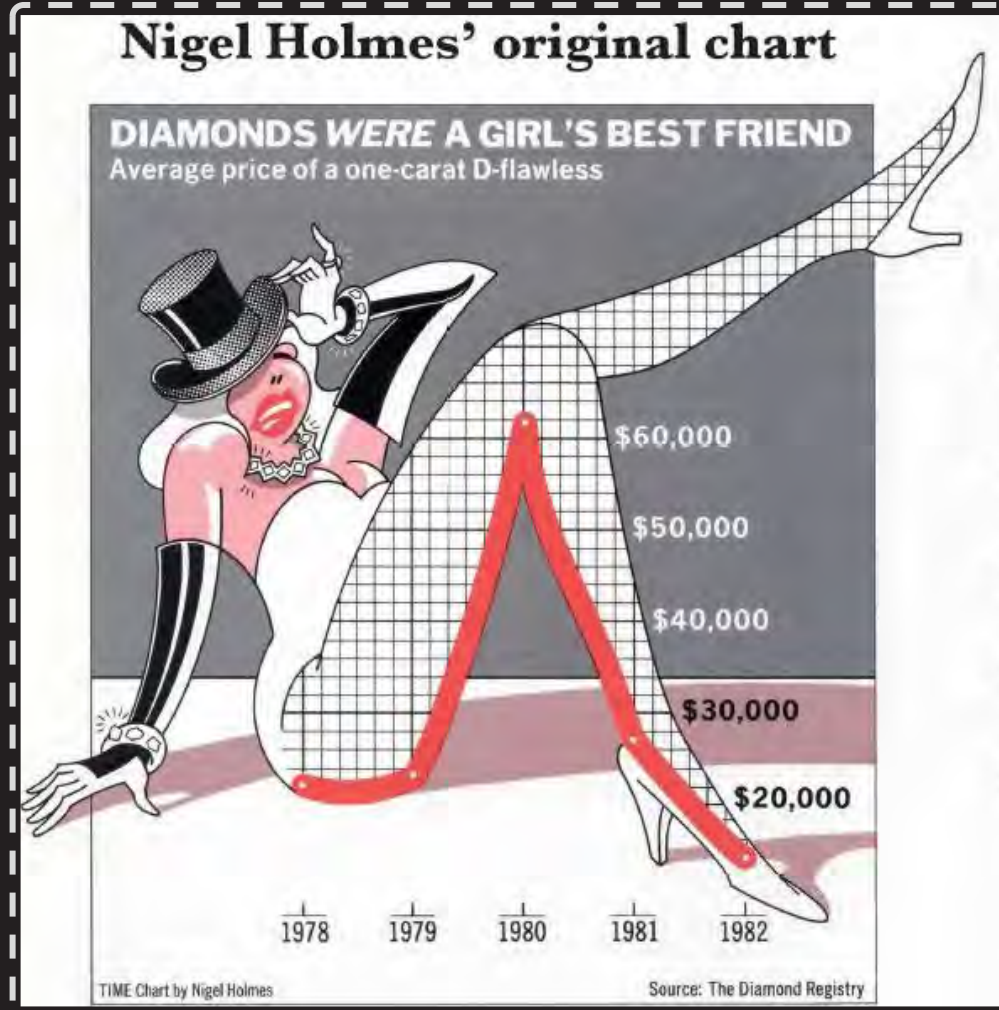
## Higher data-ink ratio



Alberto Cairo

# Nigel Holmes

Nigel Holmes is a British/American graphic designer, author, and theorist, who focuses on information graphics and information design.



# ESPRESSO AND MILK

*Let us count the ways.*

BOLD



## DOPPIO

Two shots of espresso. Straight.



## CAFFÈ AMERICANO

Shots of espresso and filtered hot water.



## LATTE MACCHIATO

Foamed whole milk marked with shots of espresso.



## FLAT WHITE

Sweet ristretto espresso shots finished with steamed whole milk.



## CAPPUCCINO

A shot of espresso topped with a deep layer of foamed milk.



## CAFFÈ LATTE

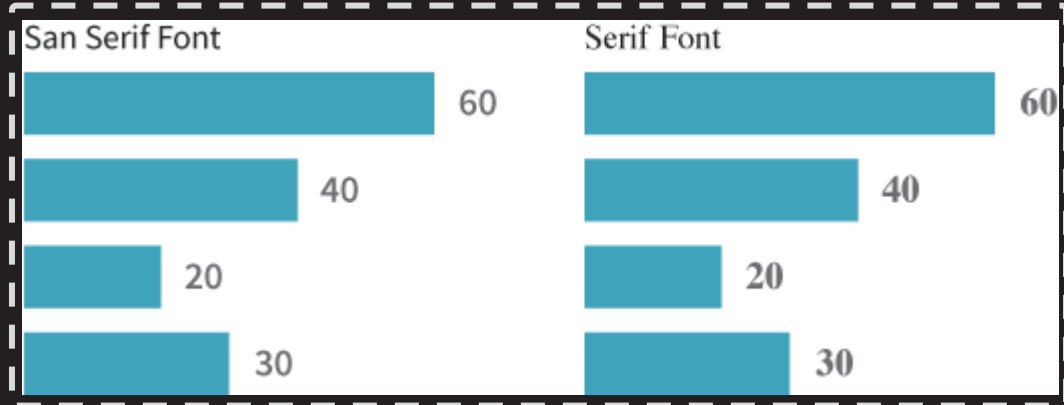
A shot of espresso in steamed milk lightly topped with foam.

CREAMY

© 2023 Zhenyuan Lu, [zhenyuanlu.com](http://zhenyuanlu.com)

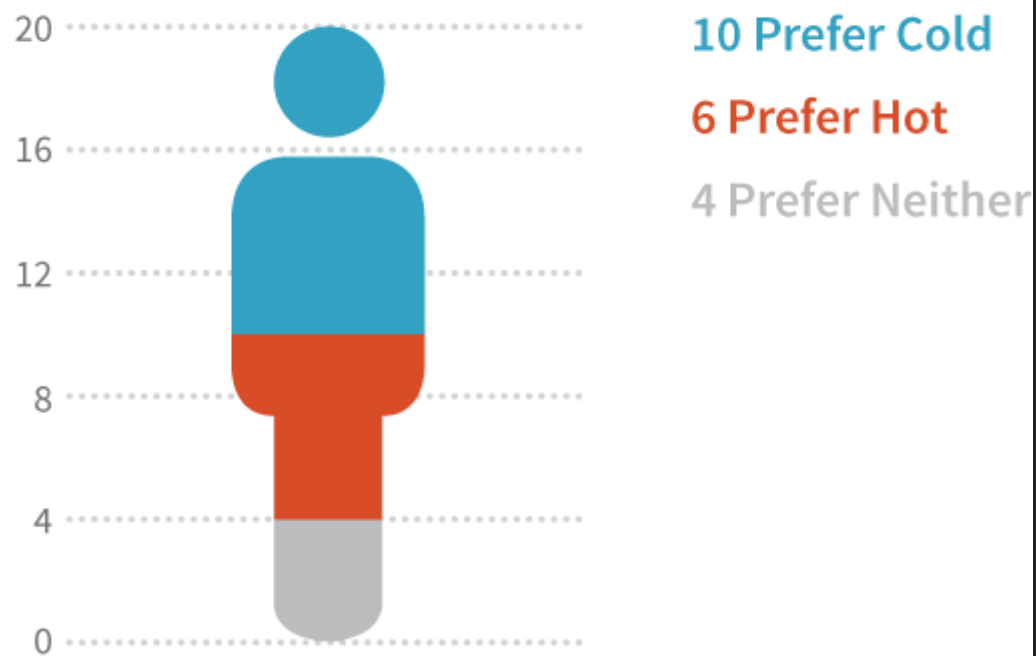
Image Credit: Starbucks

Real World Issues



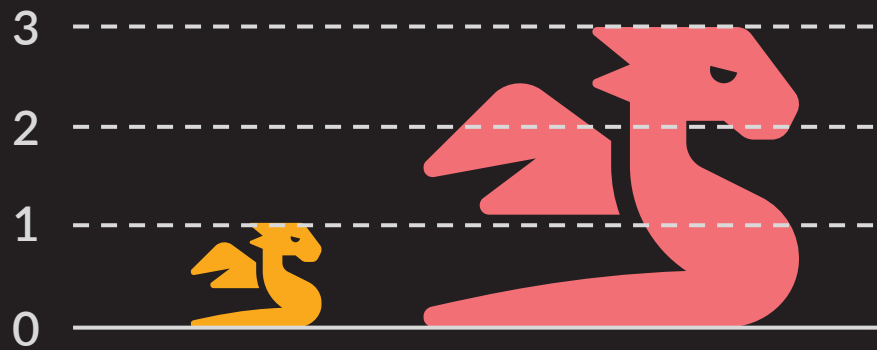
Serif

### Temperature Preference: 20 Test Subjects

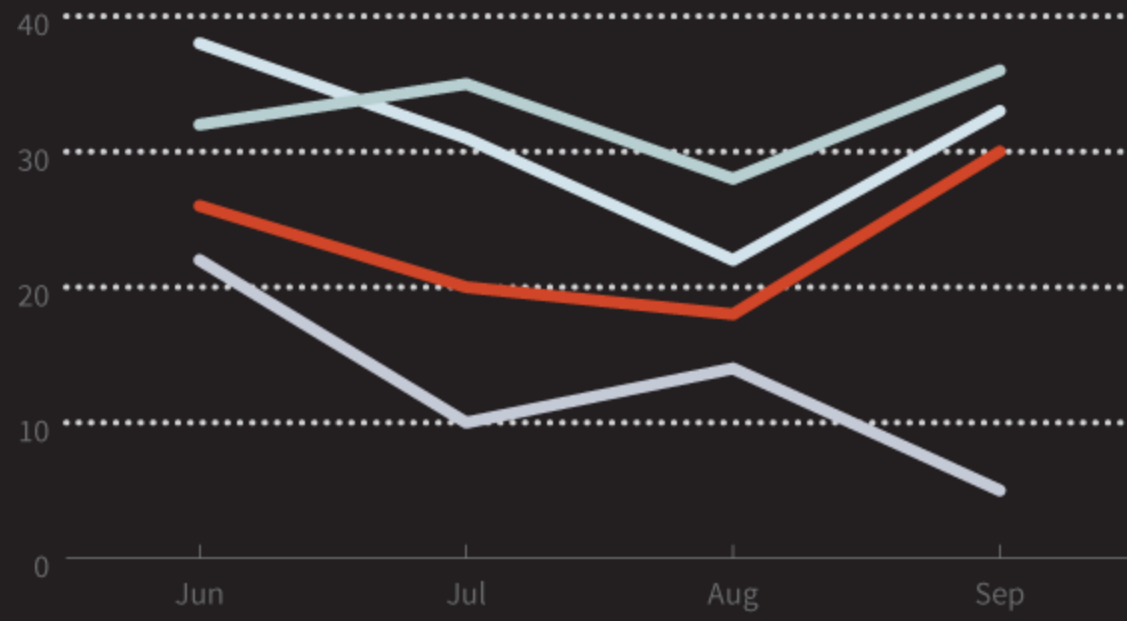


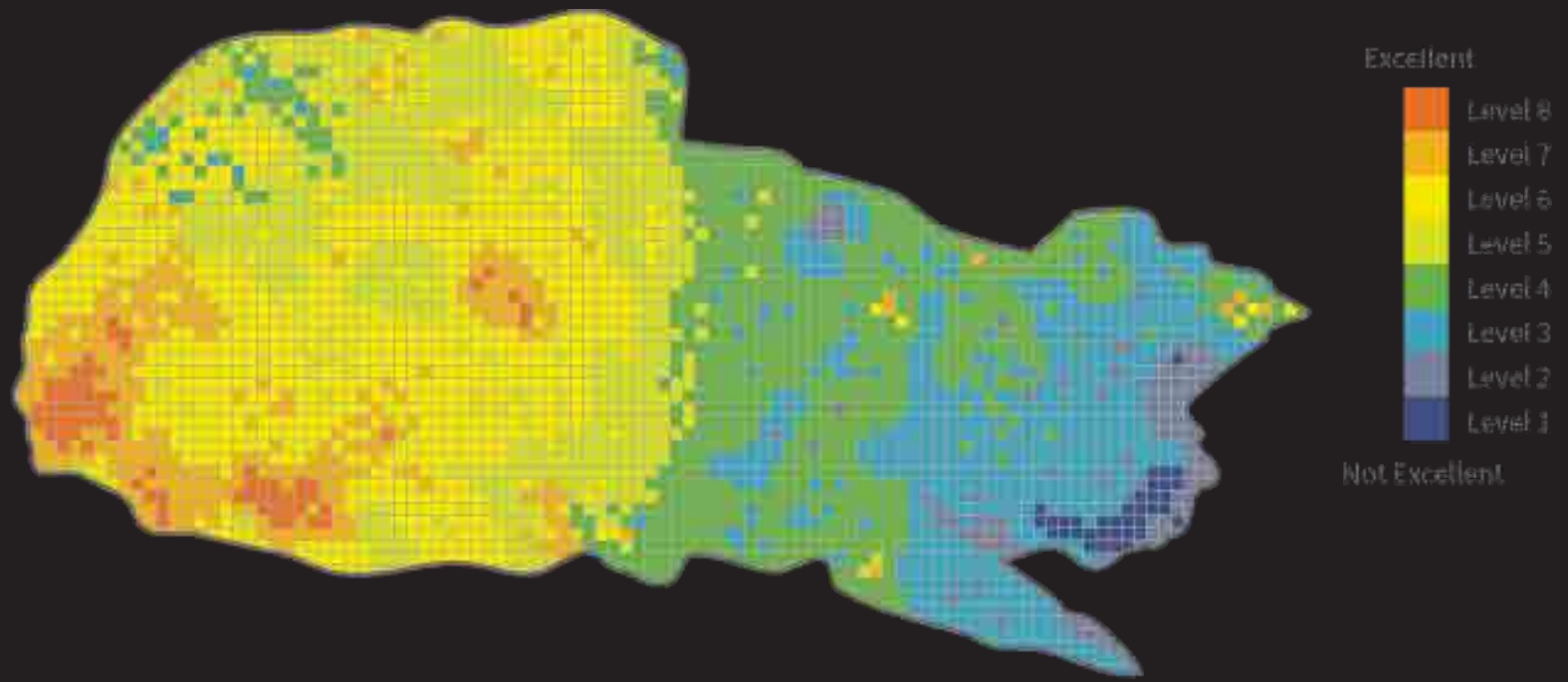
### Temperature Preference: 20 Test Subjects





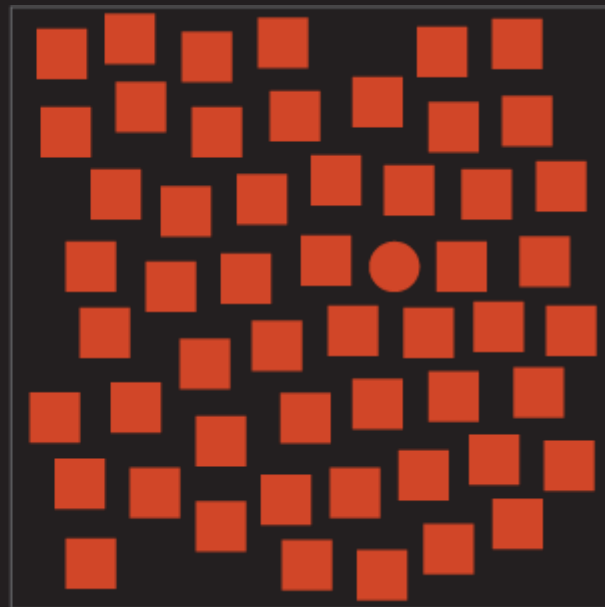
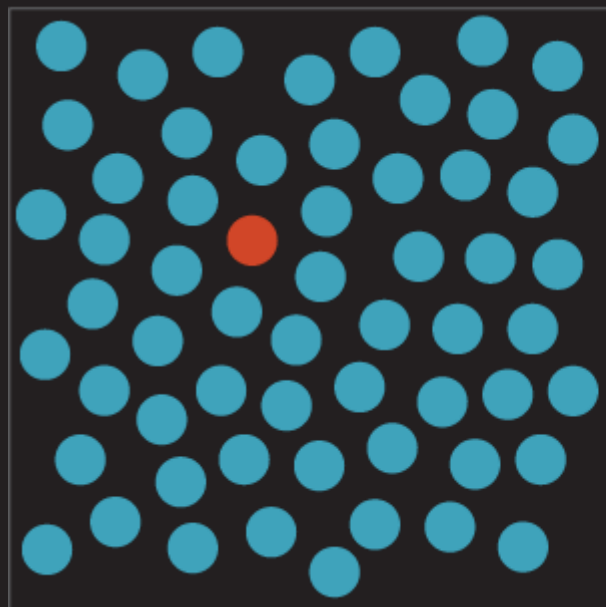


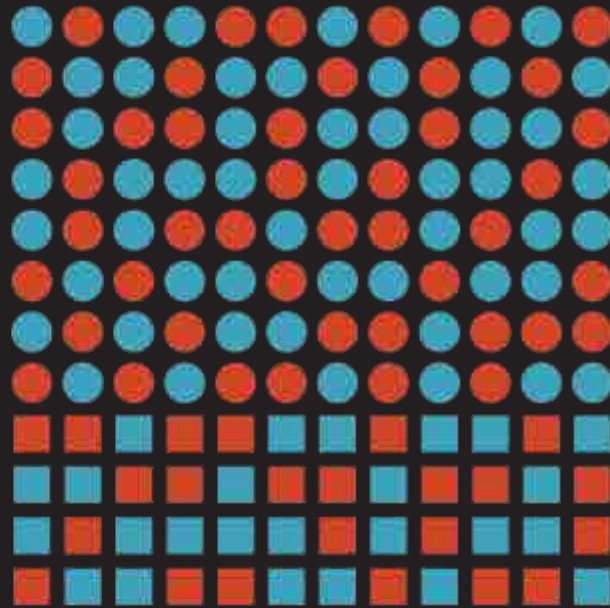
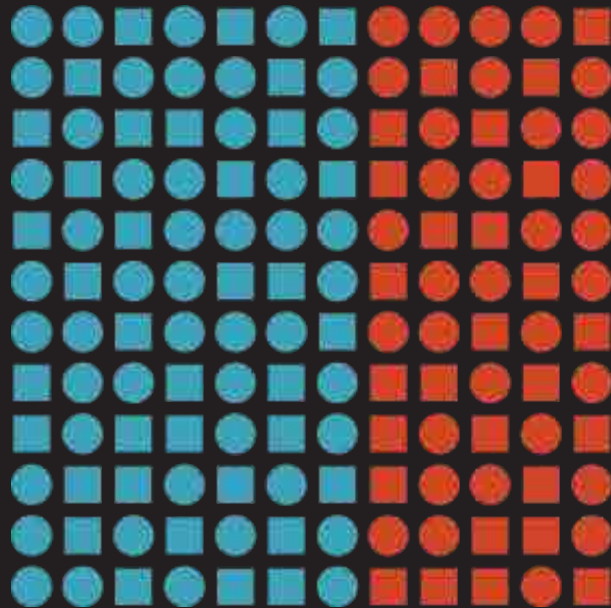




= one level difference =

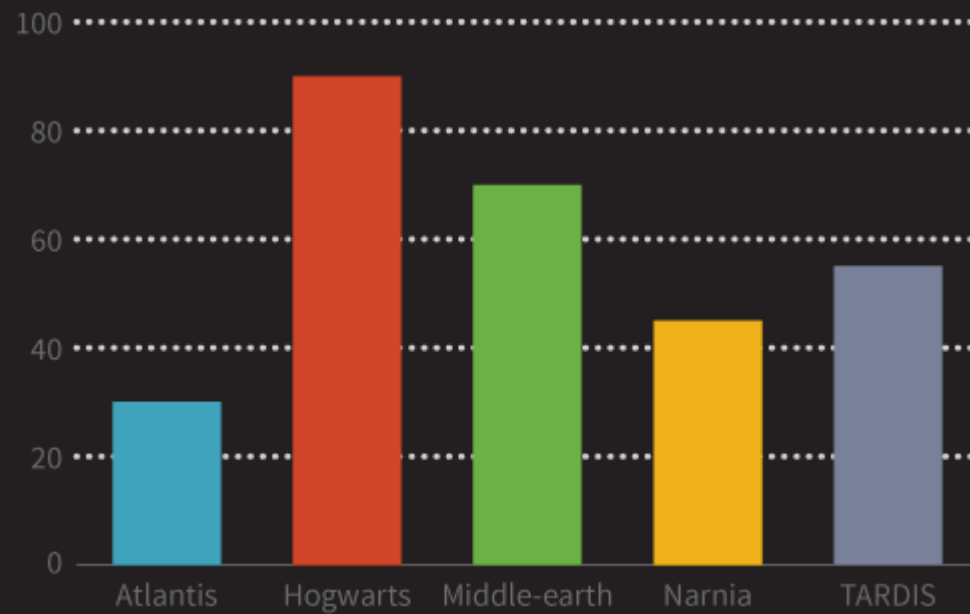
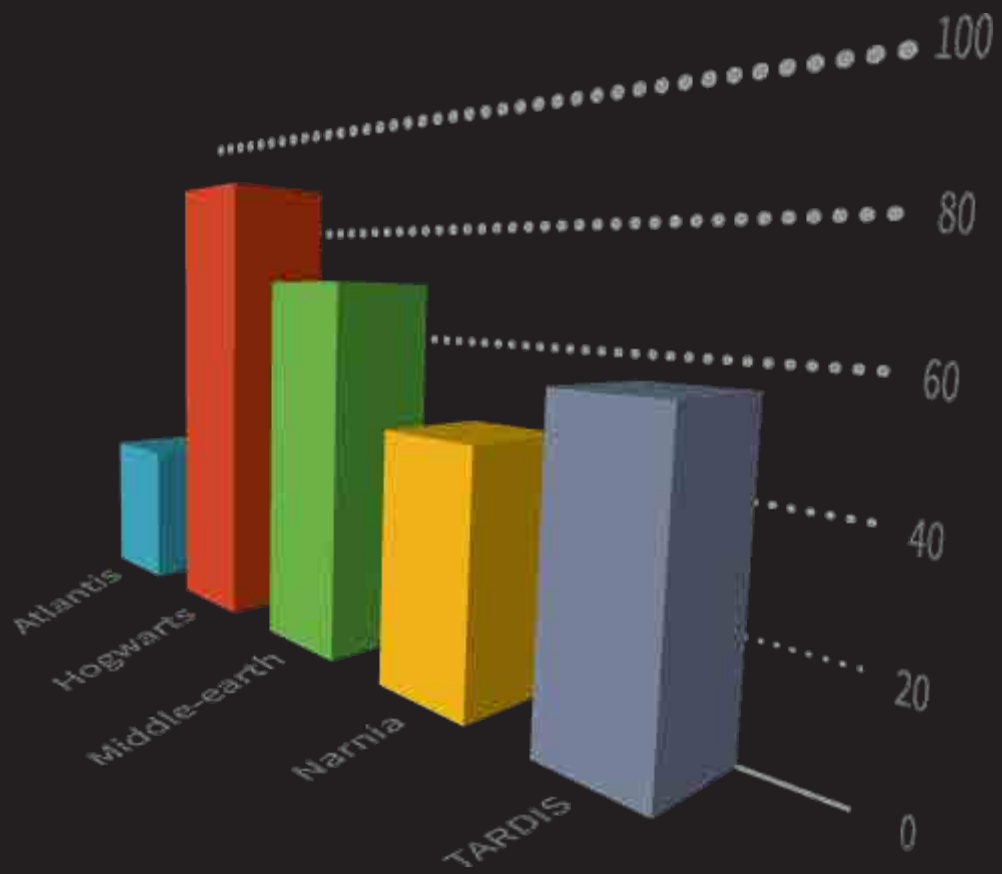






Adapted from: Healey, Christopher G., Kellogg S. Booth, and James T. Ennis. "High-Speed Visual Estimation Using Preattentive Processing." *ACM Transactions on Computer-Human Interaction* 3.2 (1996): 5.

© 2023 Zhenyuan Lu, [zhenyuanlu.com](http://zhenyuanlu.com)



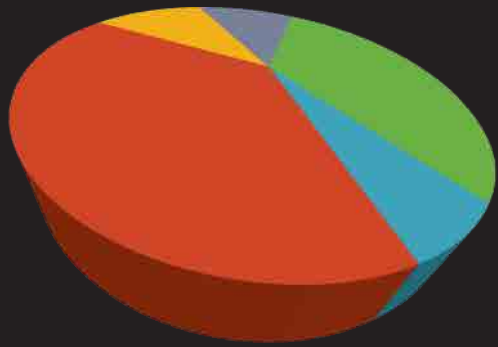
In Other Hand...

Top View

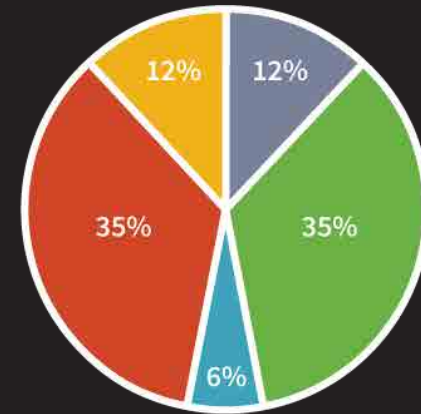


Side View

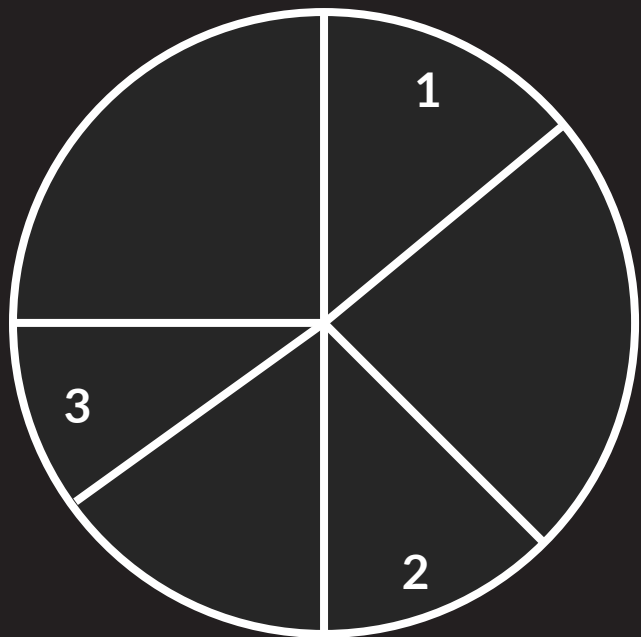




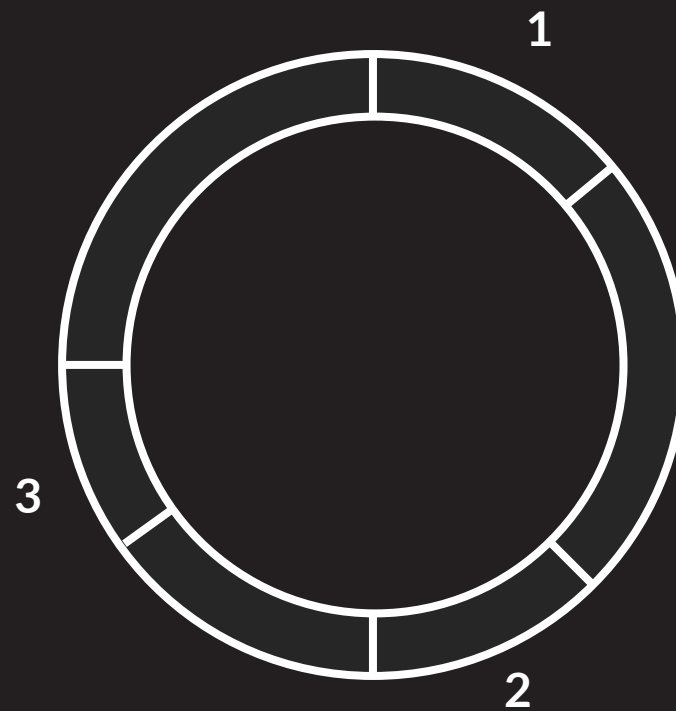
● Atlantis ● Hogwarts ● Middle-earth ● Narnia ● TARDIS



● Atlantis ● Hogwarts ● Middle-earth ● Narnia ● TARDIS

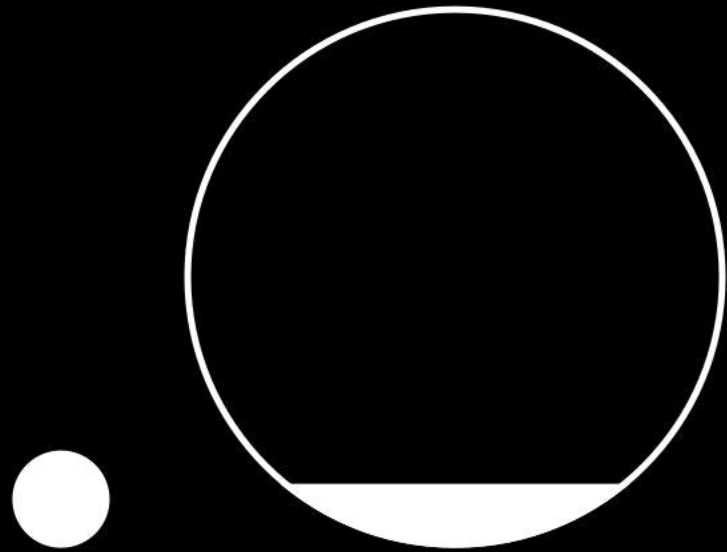


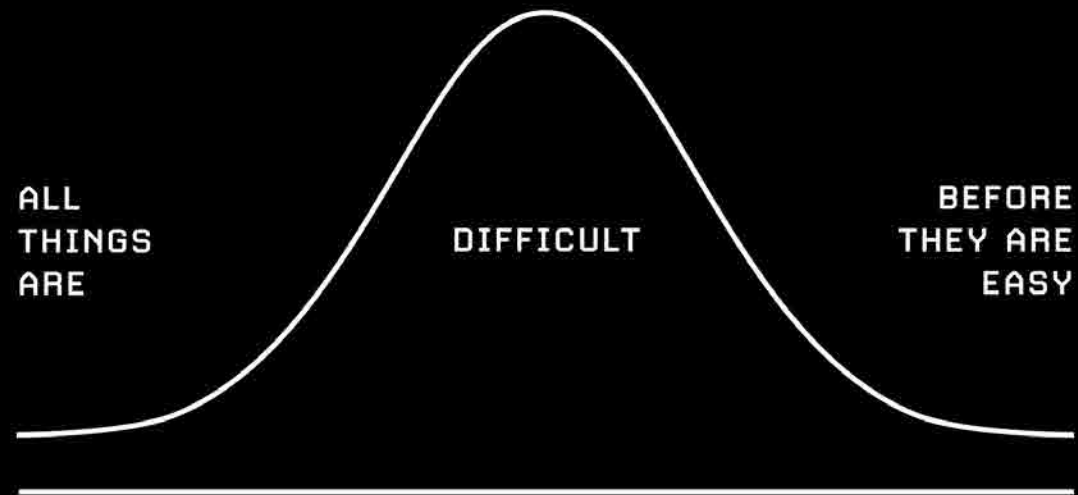
If the area 1 = area 2 = area 3 ?



Better ?

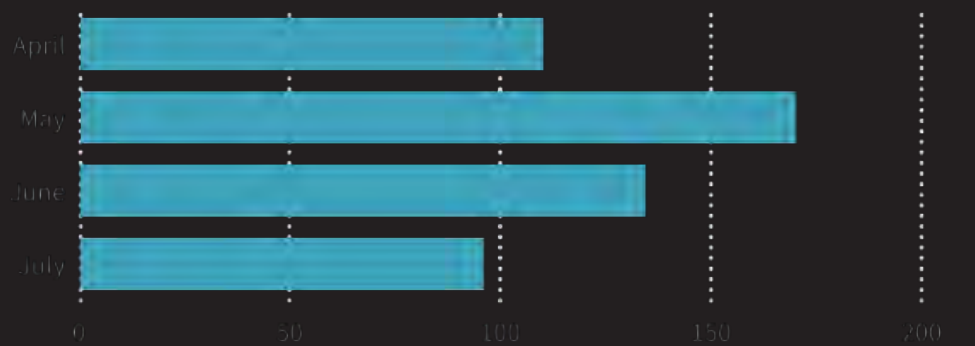
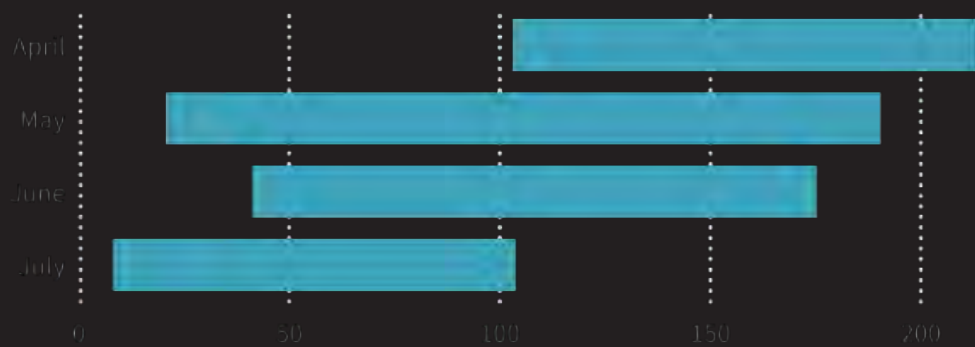




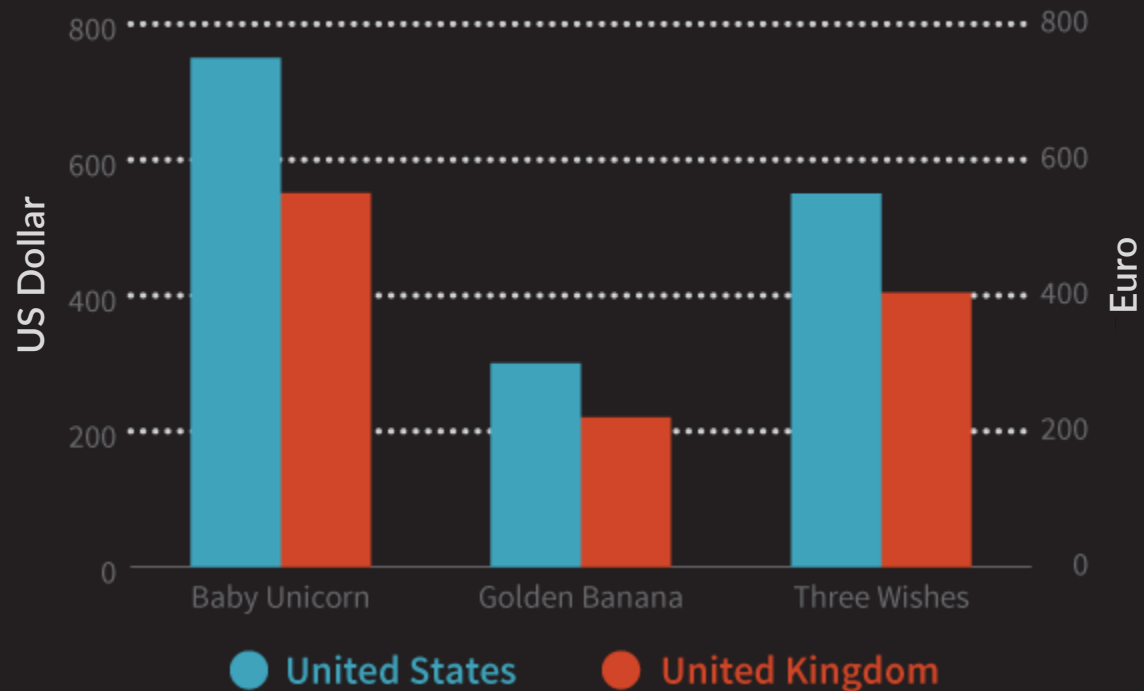


*"All things are difficult before they are easy." – Chinese Proverb*

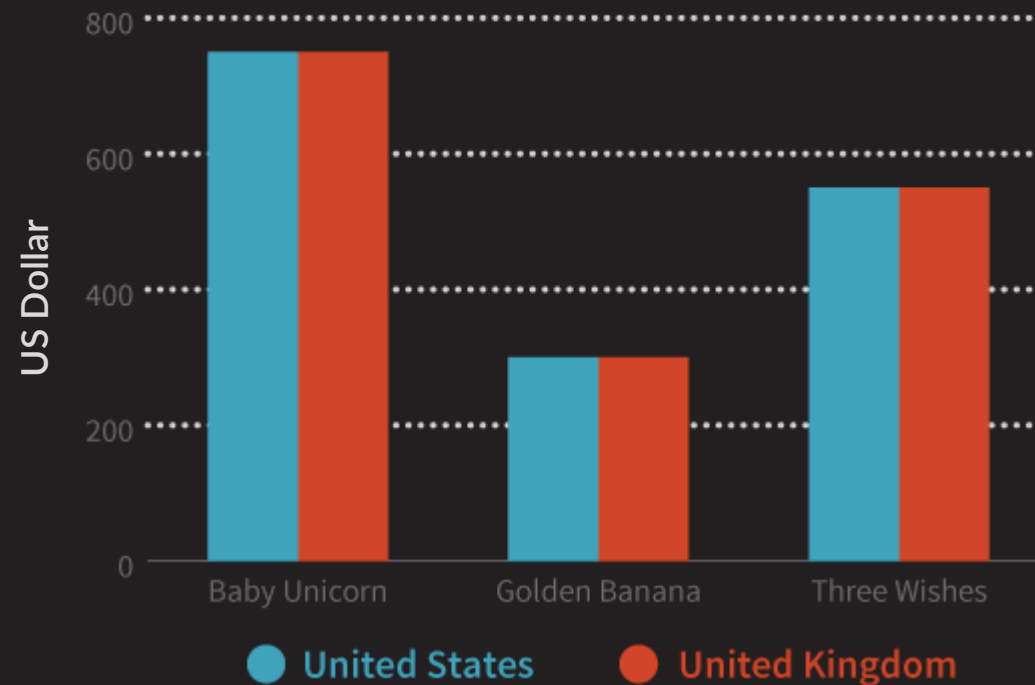
万事开头难



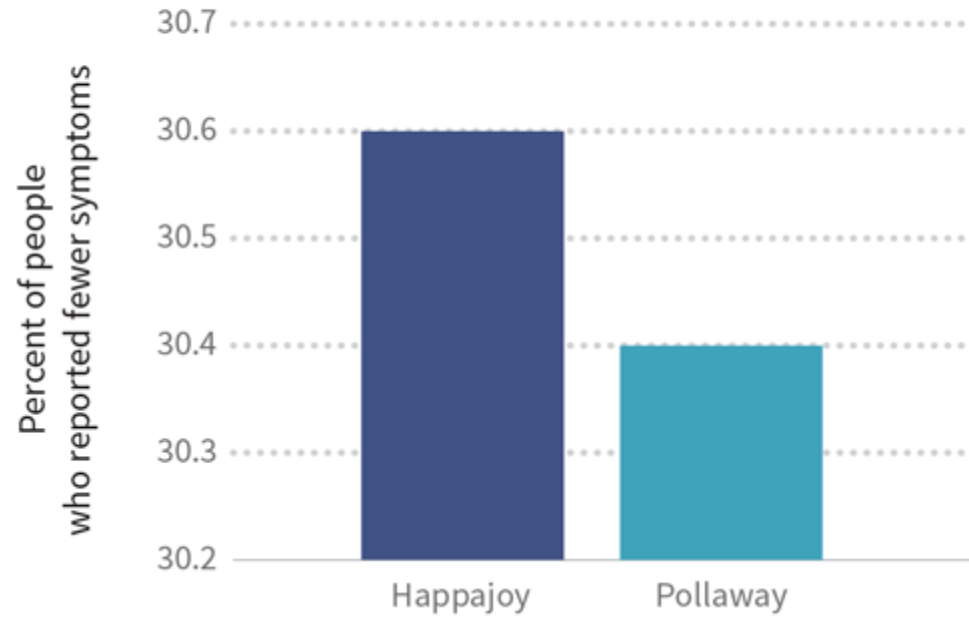
## Cost of Magical Things



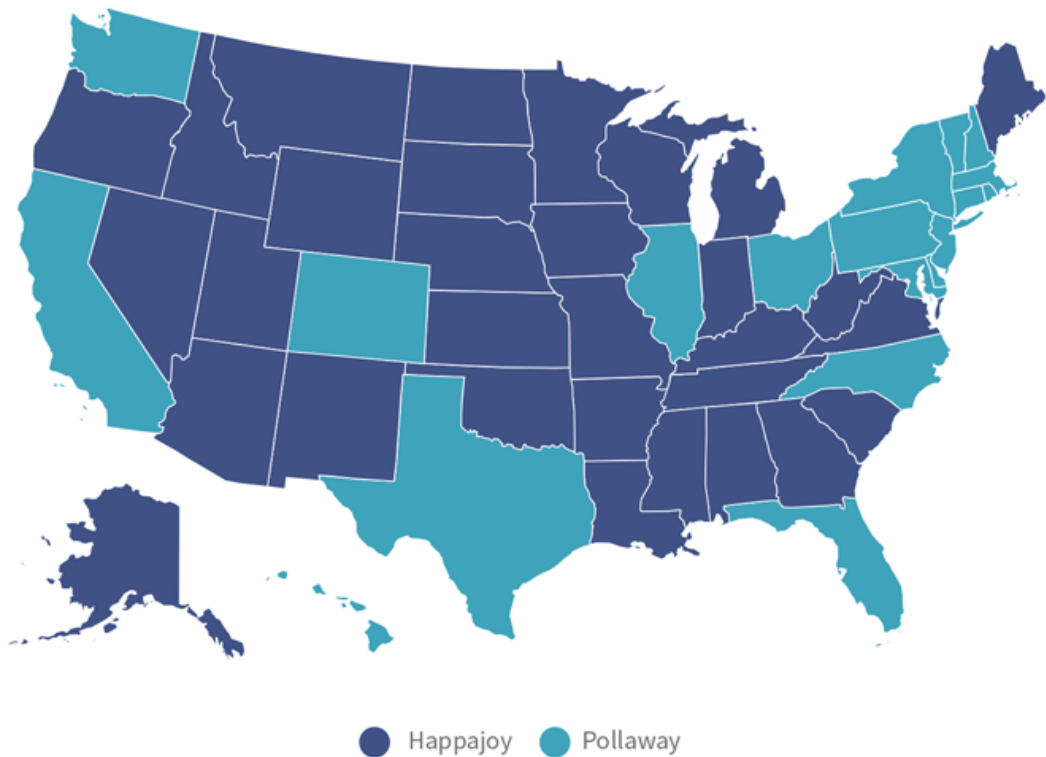
## Cost of Magical Things



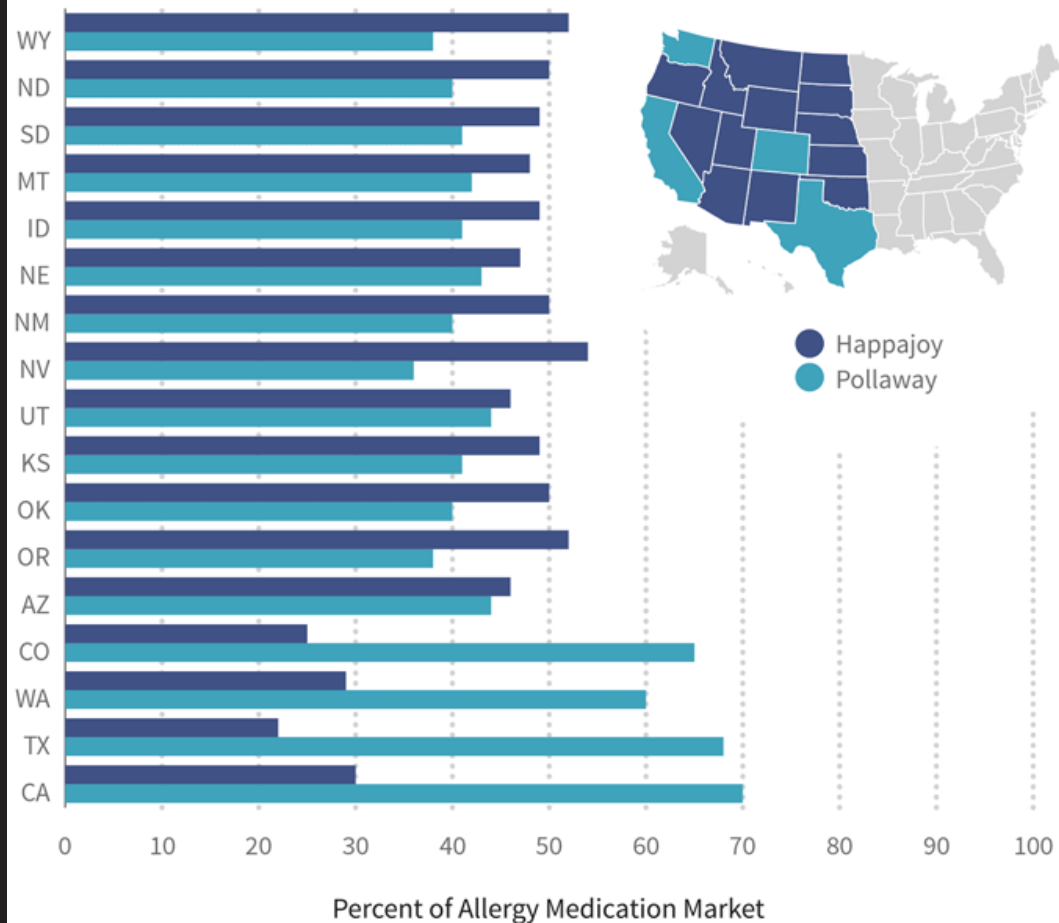
## Effectiveness of Allergy Medicines



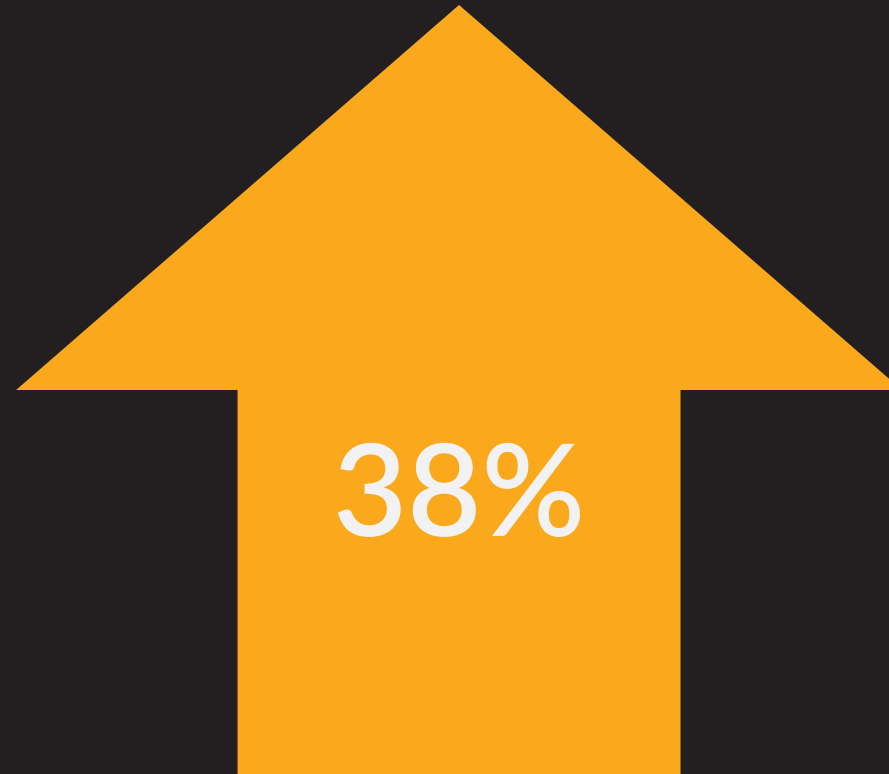
### Leading Allergy Medication in Each State



### Market Share in Each State



# 2021 Revenue



## Cumulative iPhone sales

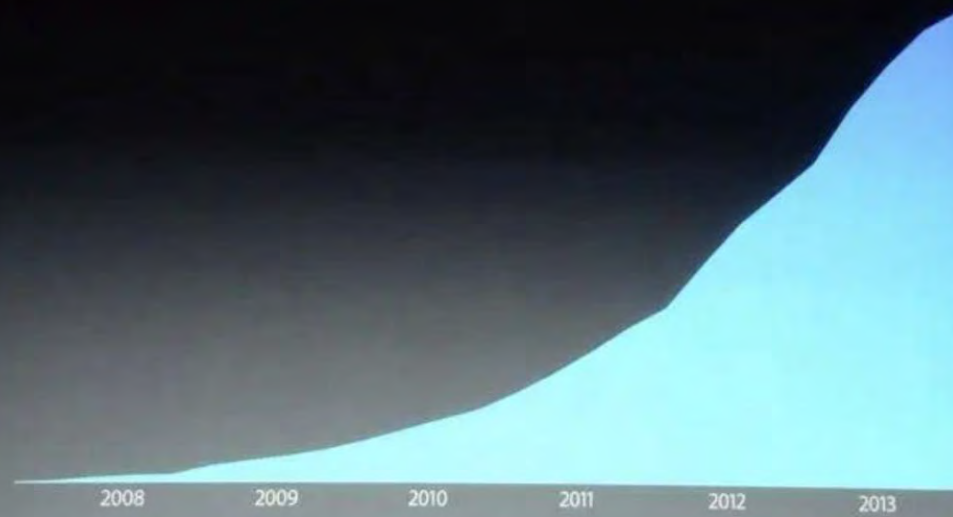
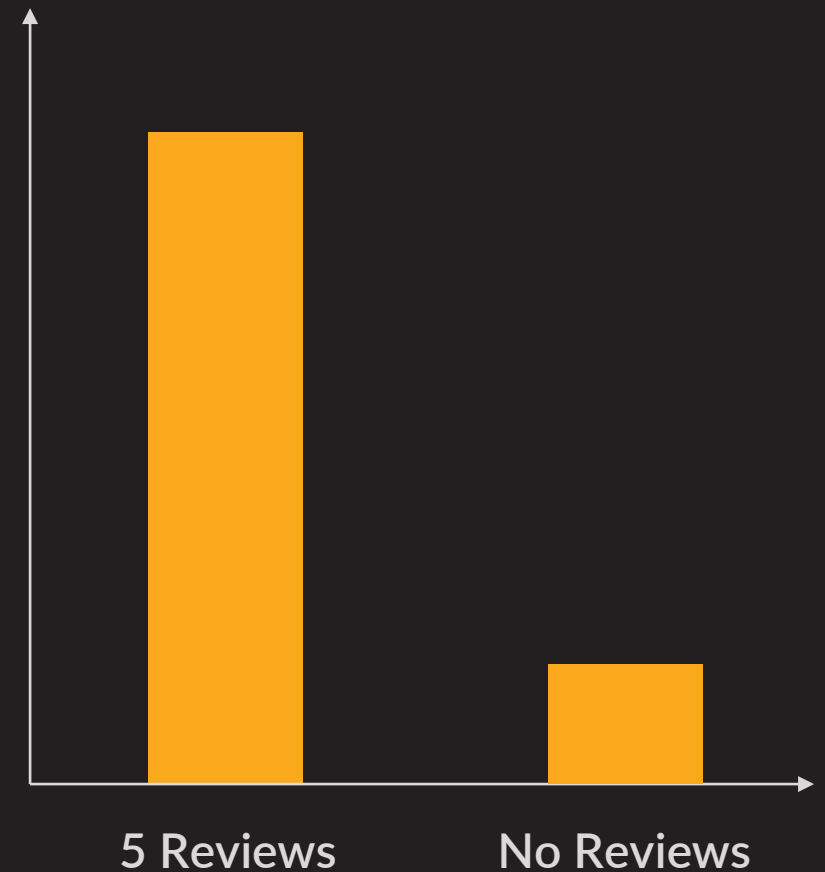


Image Credit: Apple



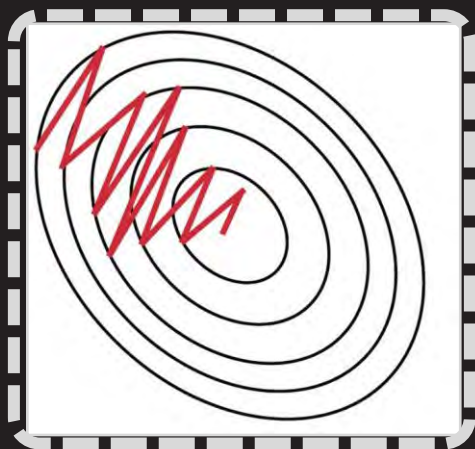
“The purchase likelihood for a product with five reviews is 270% greater than a product with zero reviews.”

Spiegel Research Center, *How Online Reviews Influence Sales*



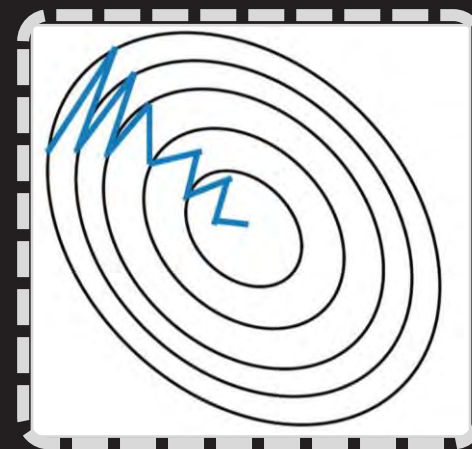
“The purchase likelihood for a product with five reviews is **270%** greater than a product with zero reviews.”

Spiegel Research Center, *How Online Reviews Influence Sales*



SGD

- For n epochs:
  - For  $\mathbf{x}, \mathbf{y} \sim D$ :
    - $\theta := \theta - \epsilon \frac{\partial l(\mathbf{x}, \mathbf{y} | \theta)}{\partial \theta}$



SGD with  
momentum

- $\mathbf{v} := 0$
- For n epochs:
  - For  $B_i$  batches:
    - $\mathbf{g} := \mathbf{E}_{\mathbf{x}, \mathbf{y} \in B_i} \frac{\partial l(\mathbf{x}, \mathbf{y} | \theta)}{\partial \theta}$
    - $\mathbf{v} := \rho \mathbf{v} + \mathbf{g}$
    - $\theta := \theta - \epsilon \mathbf{v}$









Where  $\mathbf{v}$  is the momentum term. If  $\rho$  is zero, then will return to **SGD**.

# Respect Your Audience

#F26F75

## Color Blindness Simulator

Below, you can see how #f26f75 is perceived by people affected by a color vision deficiency. This can be useful if you need to ensure your color combinations are accessible to color-blind users.

<b>Monochromacy</b>	 Achromatopsia	 Atypical Achromatopsia	
<b>Dichromacy</b>	 Protanopia	 Deuteranopia	 Tritanopia
<b>Trichromacy</b>	 Protanomaly	 Deuteranomaly	 Tritanomaly

## Take Home Notes

1. There are some basic rules
2. There is no a MUST rule
3. Accept that a graphic is a tool for understanding
4. Be curious and start making simple graphics about topics you care about.
5. Educate yourself in creating hierarchies, narratives, stories, etc., that make your presentations understandable.
6. Respect your audience.

Q & A

Vitaly Friedman (2008) "Data Visualization and Infographics", Graphics, Monday Inspiration.

Cairo, A. (2013). *The functional art an introduction to information graphics and visualization*. New Riders.

Kshitij Ranjan , Zhenyuan Lu , Shannon Weigum (2017), Development of a paper-based diagnostic platform for the detection of diarrhea causing pathogens, BMES

Zhenyuan Lu , Kshitij Ranjan , John Carrano , Jacob Carrano , Roland Schneider , Shannon Weigum (2016), Paper-based Device For Gastroenteritis Detection Integrated With Sample Preparation Cartridge, BMES

Zhenyuan Lu , Jacob Carrano , Roland Schneider , John Carrano, (2018), Point-of-care Diagnostic Cartridge Having A Programmable Fluidic Wicking Network, US Patent

Gilbert, Scott. *Developmental Biology*, 9th Edition. Sunderland, MA: Sinauer Associates Inc., 2010.

Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D.P. and Shetty, S., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography [\[link\]](#), *Nature Medicine*, Vol 25(6), pp. 954--961. Springer Science and Business Media {LLC}. DOI: [10.1038/s41591-019-0447-x](https://doi.org/10.1038/s41591-019-0447-x)

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J.J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K.C., Fauw, J.D. and Shetty, S., 2020. International evaluation of an {AI} system for breast cancer screening [\[link\]](#), *Nature*, Vol 577(7788), pp. 89--94. Springer Science and Business Media {LLC}. DOI: [10.1038/s41586-019-1799-6](https://doi.org/10.1038/s41586-019-1799-6)

MISC Credits:

David Alayón

Eleanor Lutz, 2014, <http://tabletopwhale.com/>

Hans Rosling

BBC

**DATA + DESIGN, A simple introduction to preparing and visualizing information**

© 2023 Zhenyuan Lu, [zhenyuanlu.com](http://zhenyuanlu.com)

# References