

# ICOS BIG DATA CAMP

## VISUALIZATION BASICS



---

*Visualization Basics - From Chart to Art: An Introduction to Design Science*

### General Resources –

- Visual Strategies: A Practical Guide to Graphics and Engineers, Frankel and Depace – 2012.
- Color Design Workbook: A Real World Guide to Using Color in Graphic Design, Sean Adams, 2008.
- Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization, Friendly & Denis <http://euclid.psych.yorku.ca/SCS/Gallery/milestone/>
- Graphical Perception: Theory, Experimentation, and Application to the Development of
- Graphical Methods. Cleveland & McGill 1984.
- Articles found within <https://www.pictureasportal.com/pages/resources>

# Visualization of Information -

How can we:

*1) convey our data most informatively (accuracy)?*

*and...use the tools we have, to*

*2) create interesting and approachable graphics (aesthetics)?*

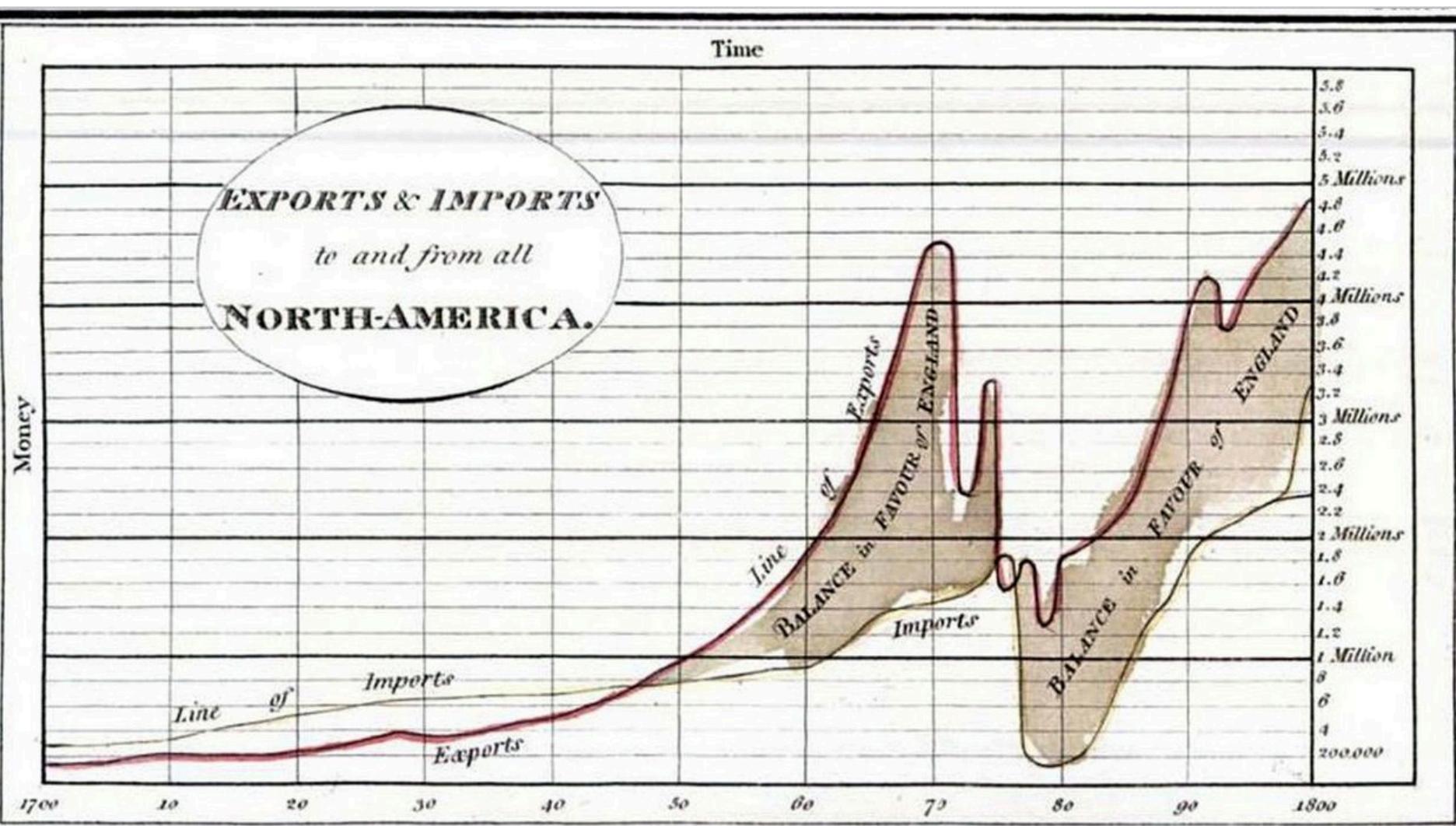
# The beginning of a more familiar statistical data visualization

With the beginning of modern statistics & probability theory in (late 1700s – 1900s) came a surge of info-graphics and visualizations –

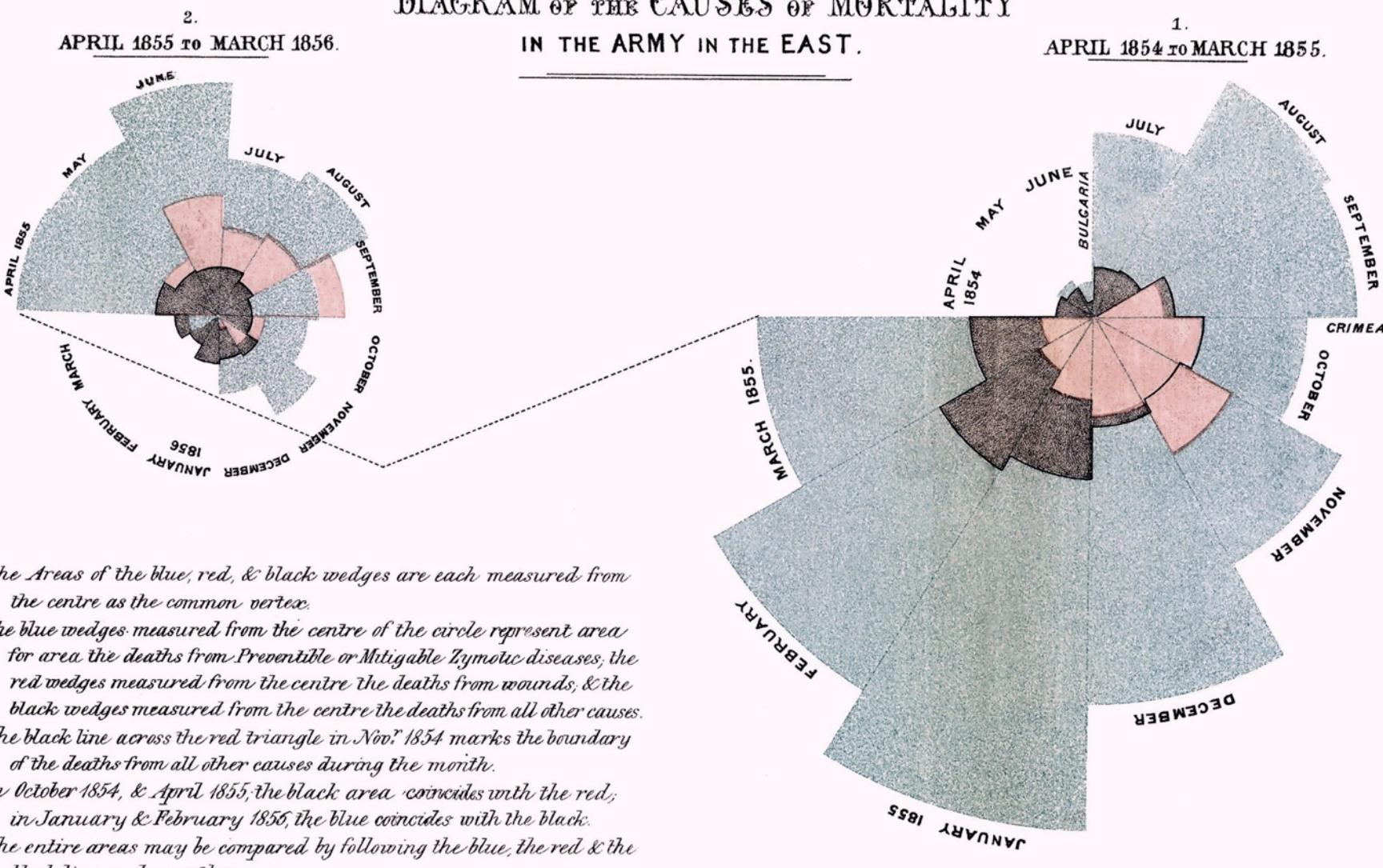
Let's take a look at some of these early charts -

- 1) What these graphics are conveying?
- 2) What you find familiar about them / unfamiliar?
- 3) What do you find confusing, what do you find interesting?

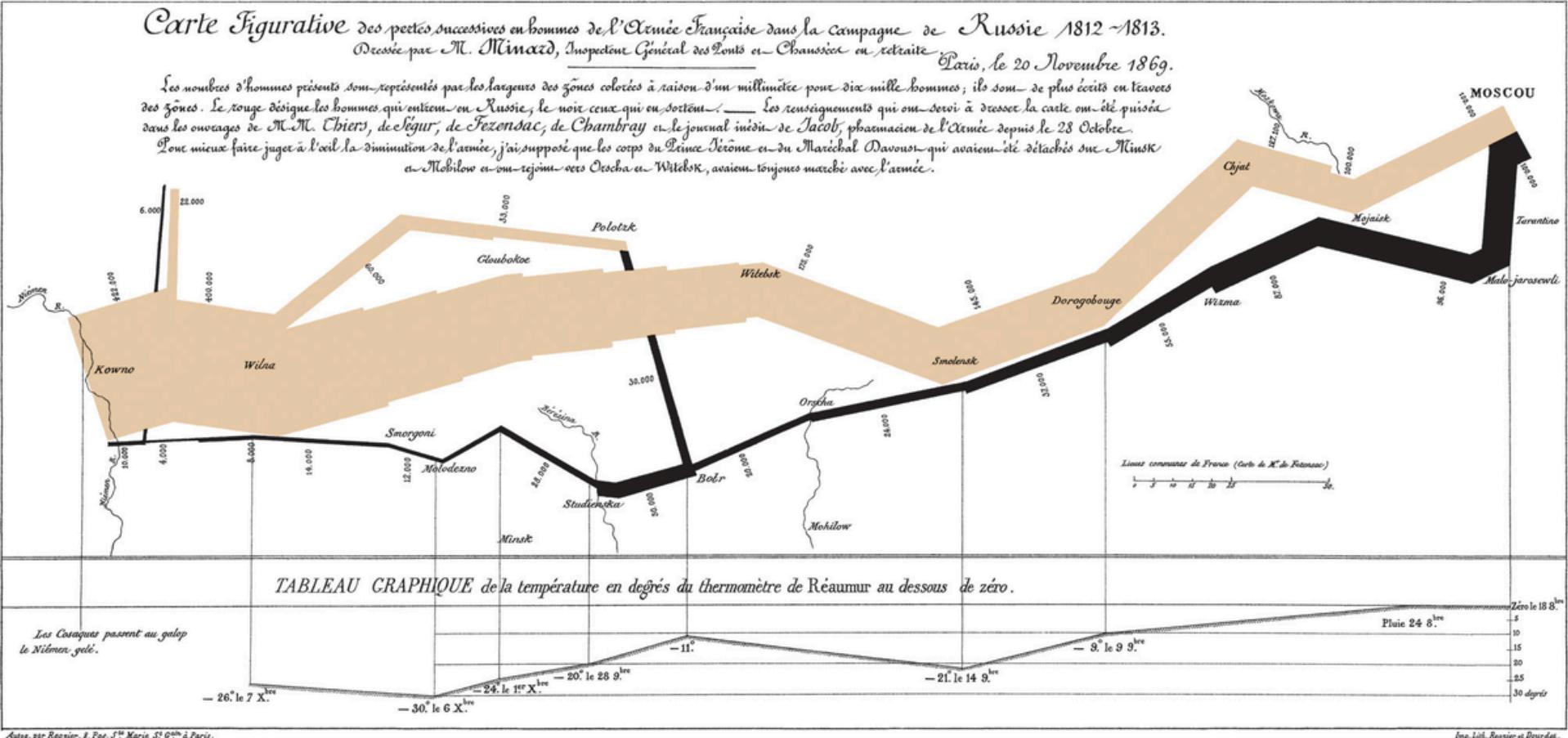
# William Playfair's Charts - 1786



# Florence Nightingale's Report - ~1858



# The “Minard Map” - 1869



# What did these all have in common?

- Why do you think they were made?

# What did these all have in common?

- Why do you think they were made?
  - The main point of these data graphics were to convince someone (mostly a government) of certain ideas or realities to push forward public policy
- Their choice of presentation of the data is therefore made to create conversation around a certain topic or point -

# What did these all have in common?

- Why do you think they were made?
  - The main point of these data graphics were to convince someone (mostly a government) of certain new ideas or realities
- Their choice of presentation of the data is therefore made to create conversation around a certain topic or point –

How can we implement “good design” as we make our own visualizations and infographics for research?

# What constitutes a “good visualization”?

- This is a rather subjective question to ask people – what do YOU think?

# What constitutes a “good visualization”?

- This is a rather subjective question to ask people – what do YOU think?
  - Does color matter?

# What constitutes a “good visualization”?

- This is a rather subjective question to ask people – what do YOU think?
  - Does color matter?
  - What colors are best?

# What constitutes a “good visualization”?

- This is a rather subjective question to ask people – what do YOU think?
  - Does color matter?
  - What colors are best?
  - Is 3D better than 2D?

# What constitutes a “good visualization”?

- This is a rather subjective question to ask people – what do YOU think?
  - Does color matter?
  - What colors are best?
  - Is 3D better than 2D?
  - What about interactive graphics? Are those better?

# What constitutes a “good visualization”?

- This is a rather subjective question to ask people – what do YOU think?
  - Does color matter?
  - What colors are best?
  - Is 3D better than 2D?
  - What about interactive graphics? Are those better?

What it all comes down to is: – who is your audience and what point are you trying to make

As you make your graphics – ask yourself the following questions...

# First Step Making a Visualization – Ask Yourself:

- 1) Is the graphic explanatory or exploratory?
  - **Explanatory** – you are trying to make a point, ask yourself what is the point you are trying to make?
  - **Exploratory** – you are inviting the viewer to explore the dataset, make their own inferences

Often the very first graphics you make in your work and research will be exploratory and then you finalize into an explanatory graphic.

# First Step Making a Visualization – Ask Yourself:

- 2) How will the graphic be used?
  - In an oral presentation you want something understood without much staring at the data – simplicity is best in this case
  - In a paper or a written document – perhaps more detail that invites the viewer to explore is better.

Above all – know your audience – what can you take for granted they know, what might they need as explanation

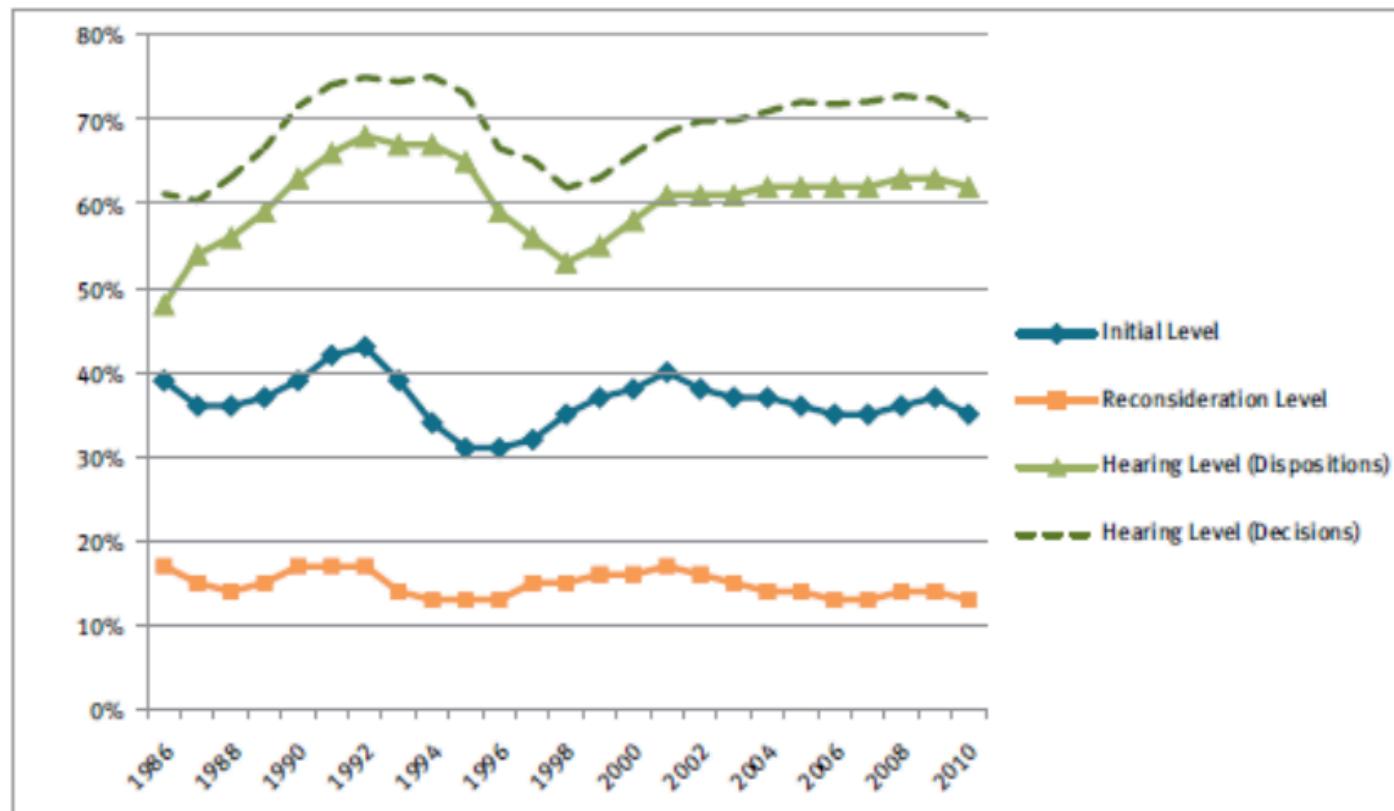
# First Step Making a Visualization – Ask Yourself:

- 3) What is the first thing you want the audience to see?
- All graphics have a “flow” to them. Most audiences are familiar with bar charts, line plots etc – if you want to get creative try to make the viewer know where to look and what the message should be
  - You can use annotation, lines etc.
  - Just try to keep it simple.
- For example...

# Example of accentuating communication

## Integrate Text and Graphics

### 7. Combined DI and SSI Allowance Rates at Each Level of Adjudication—Fiscal Years 1986-2010



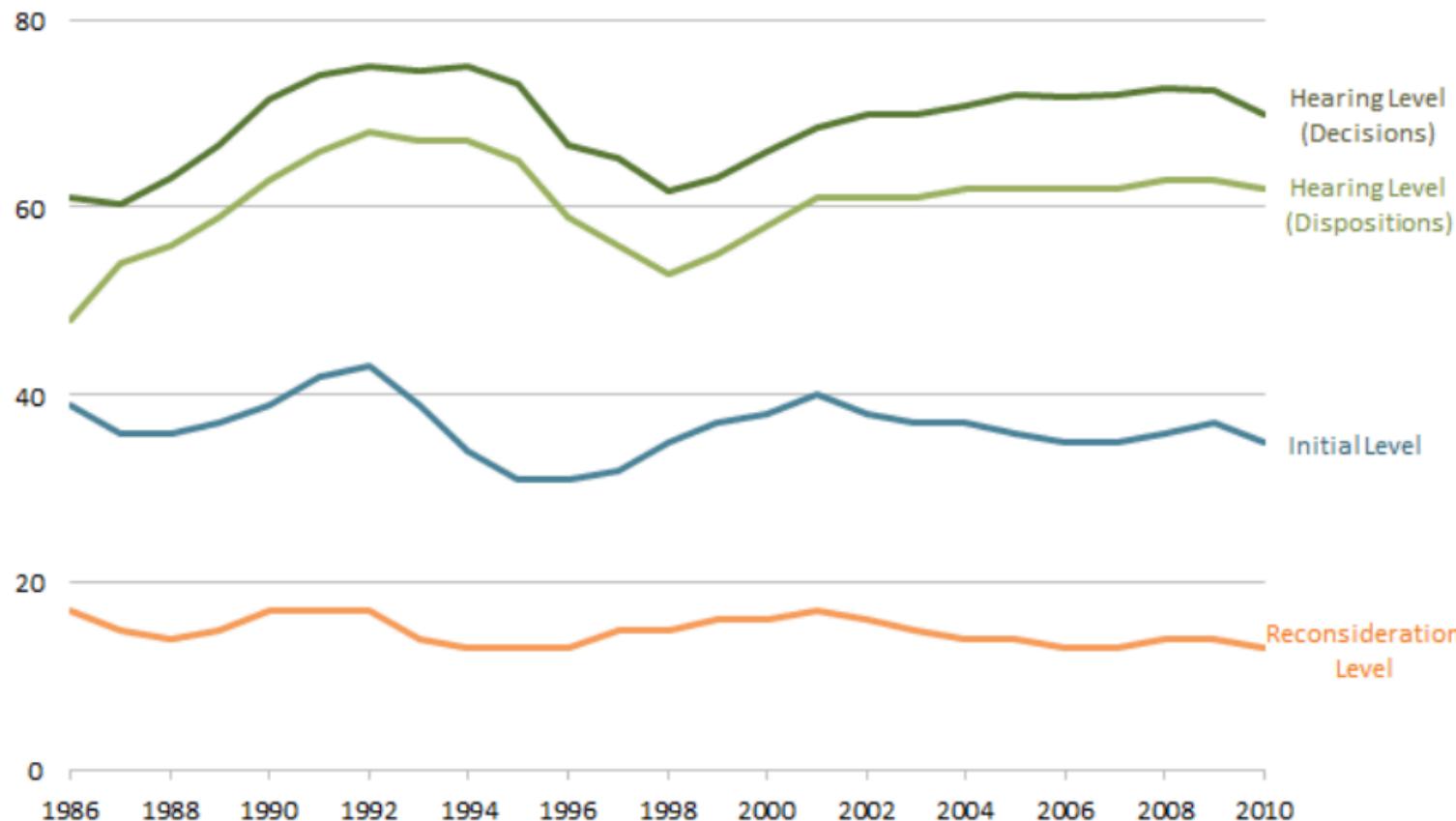
Social Security Advisory Board, *Aspects of Disability Decision Making: Data and Materials*, February 2012

Adapted from Telling Visual Stories About Data, Congressional Budget Office, Fontaine, 2014.

# Example of accentuating communication

## Integrate Text and Graphics

7. Combined DI and SSI Allowance Rates at Each Level of Adjudication—Fiscal years 1986–2010



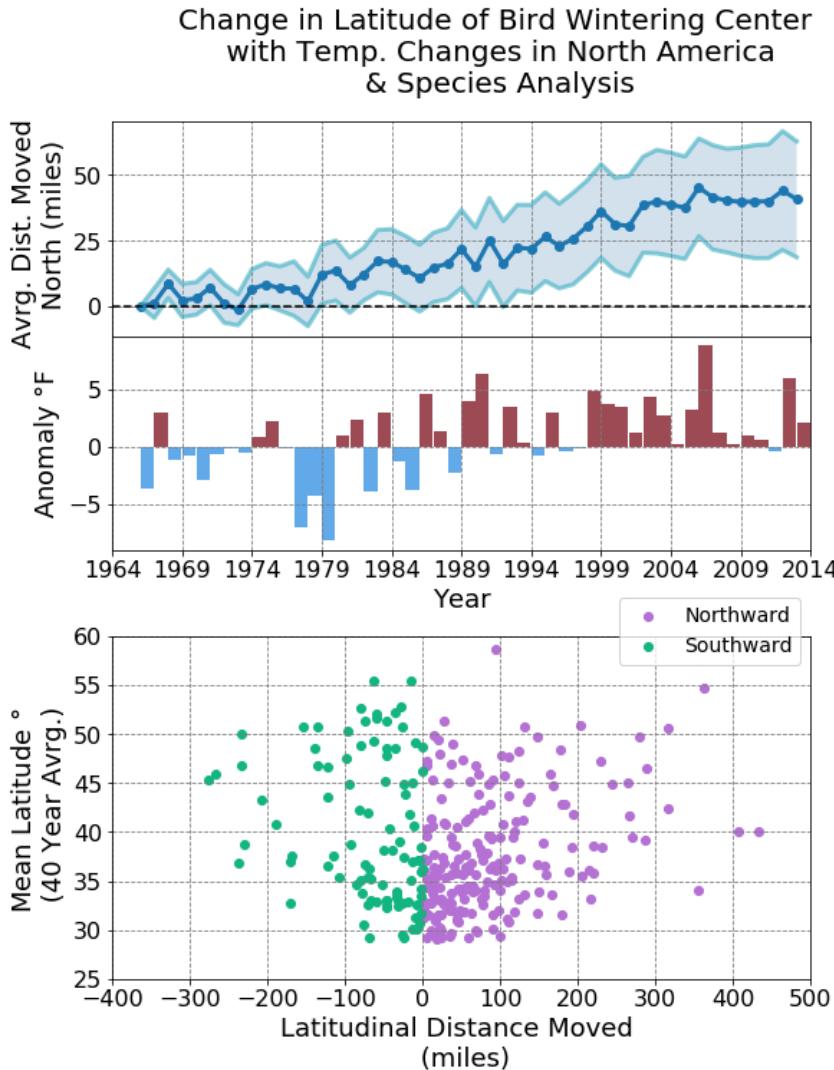
Adapted from Telling Visual Stories About Data, Congressional Budget Office, Fontaine, 2014.

# Once you have an idea of what you want to present and to whom...

*Consider the following steps and design elements*

- Compose
  - Organizing elements, defining relationships
- Abstract
  - Define and represent the meaning
- Color (my personal favorite)
  - Chose your colors to highlight relationships, label, attract or downplay interest
- Layer
  - Overlap multiple objects to create relationships
- Refine
  - Edit and simplify to the most direct communication possible

# Element 1 – Composition – how your figure “flows” *What relationships are you highlighting?*

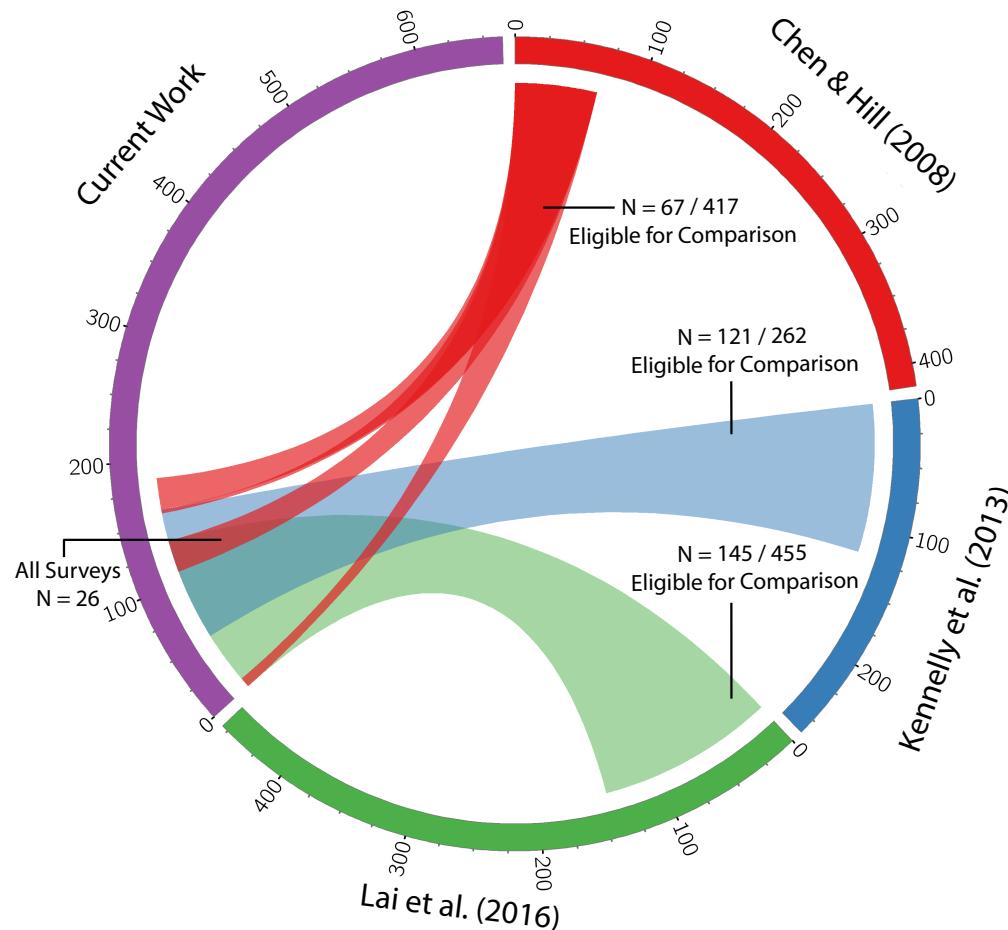


This is an example of data we are going to use in Python -

We merged this axis, to make them more comparable for the reader

While these data are of different time (non time series) so we keep them separate

# Element 1 – Composition – how your figure “flows” *What relationships are you highlighting?*



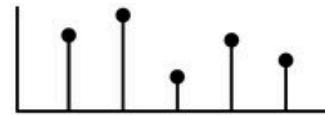
You can also have a single figure display a flow

This is a relational diagram, made with a program called Circos – here the flow is first circular then across as designated by the connection lines.

Source: [Azari et al., 2018.](#)

accurate judgement

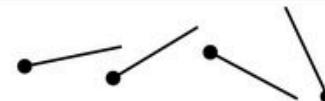
position



length



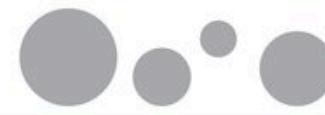
direction



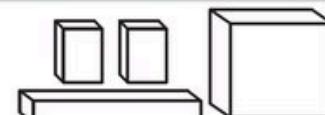
angle



area



volume



curvature



lightness



color saturation



color hue



generic judgement

## Element 2 – Abstract

What type of abstraction is supposed to represent your data? Define and representation.

*Some general guidelines from hallmark graphics study on accuracy perception by Cleveland & McGill*

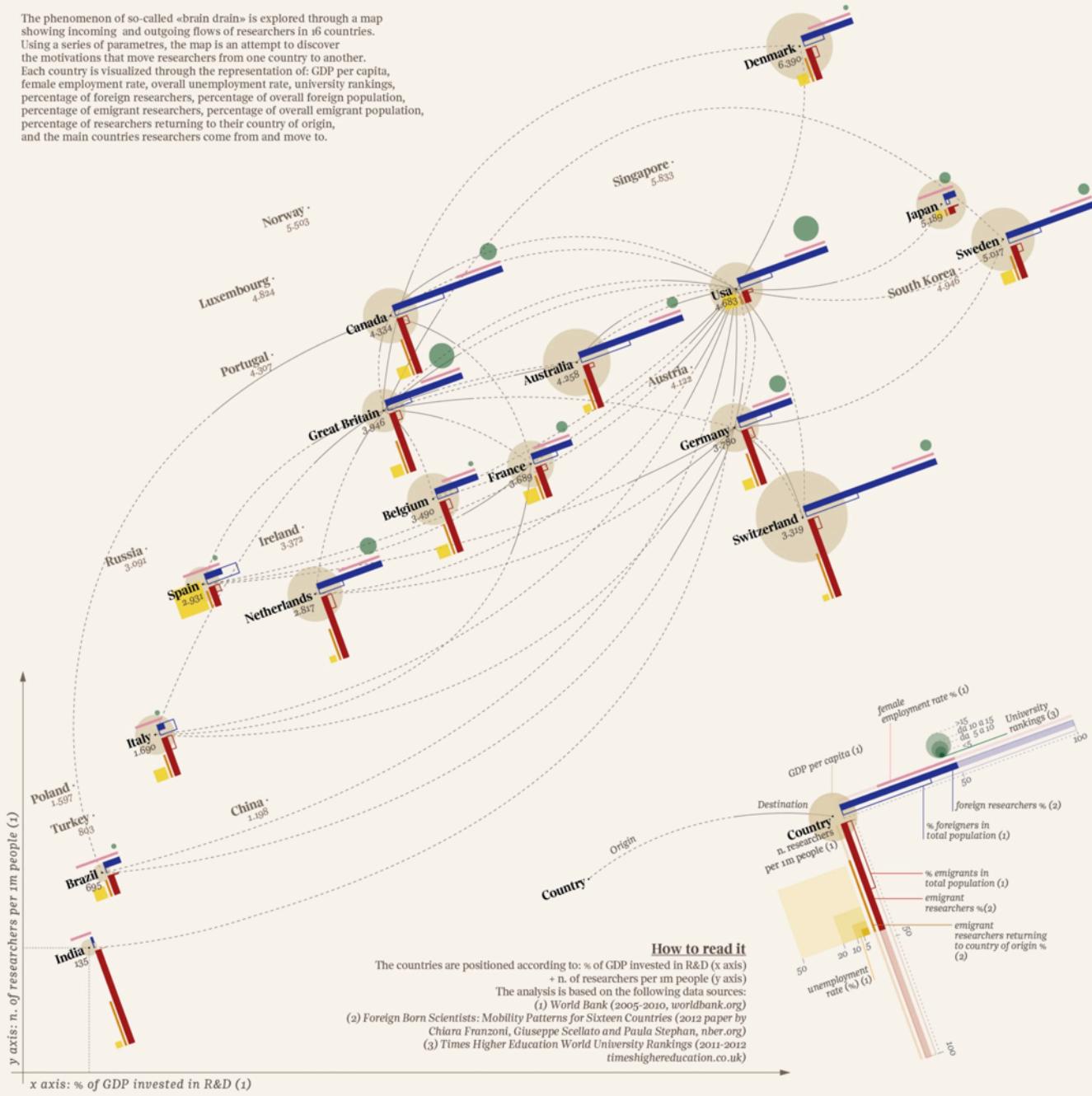
Source: Adapted from Cleveland & McGill - From  
<https://www.gabrielaplucinska.com/blog/2017/8/7/pie-charts>

# Element 2 – Abstract

You can get rather creative here – for example –

This chart is creating a new style of abstraction to communicate “brain drain” between 16 countries

The phenomenon of so-called «brain drain» is explored through a map showing incoming and outgoing flows of researchers in 16 countries. Using a series of parameters, the map is an attempt to discover the motivations that move researchers from one country to another. Each country is visualized through the representation of: GDP per capita, female employment rate, overall unemployment rate, university rankings, percentage of foreign researchers, percentage of overall foreign population, percentage of emigrant researchers, percentage of overall emigrant population, percentage of researchers returning to their country of origin, and the main countries researchers come from and move to.

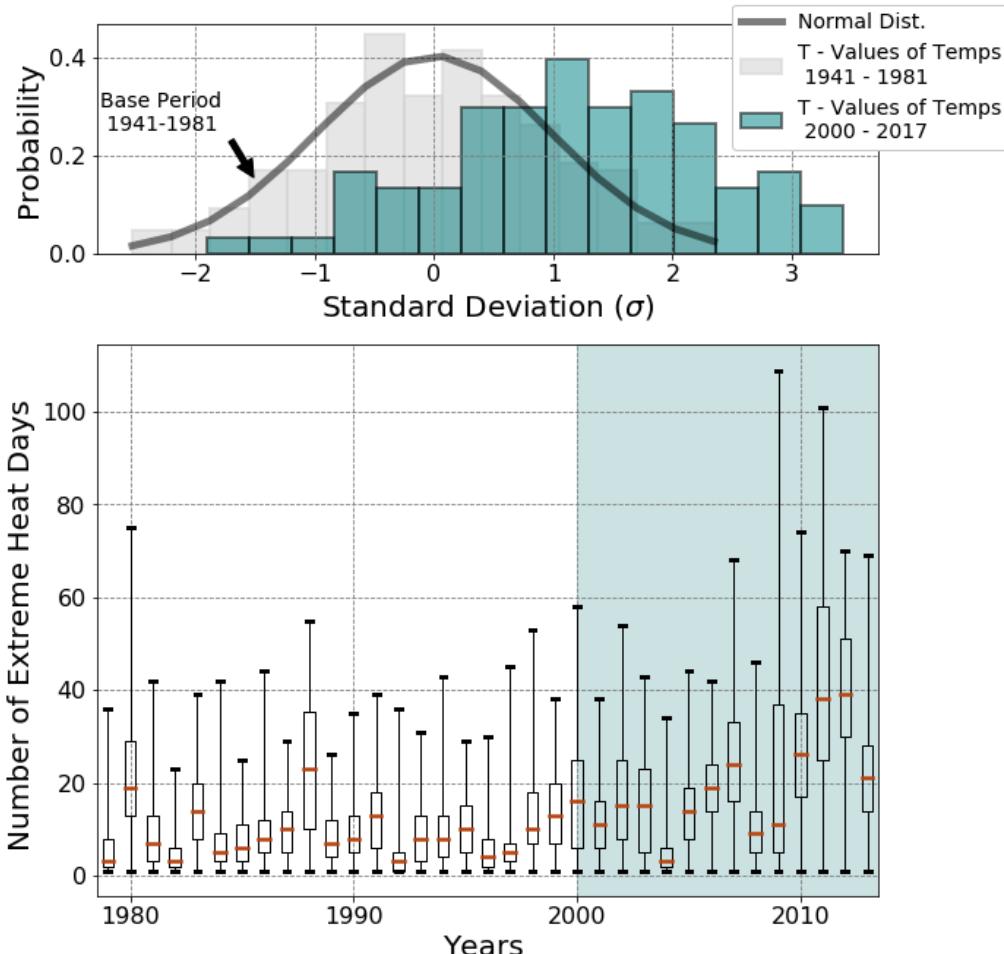


Source: The global Brain Drain—Accurat for La Lettura, Corriere della Sera

# Element 3 – Color

Use color to highlight, to show relationships, to indicate values

*“colors are the mother tongue of the subconscious” – Carl Jung*



We spend a lot of our time with colors in making visualizations

Here we tied together two times periods by signifying color

Figure from <https://github.com/astro-abby/>  
Based on Frankel and DePace, 2012.

# Element 3 – Color

## Some guidelines of color in design

Consider Conveying  
Information Or Meaning By  
Color Choice –

- Red usually will highlight something extreme or something “hot”
- In general if you are graphing a location - green is good for land, blue for ocean etc



Old Guitarist – Pablo Picasso

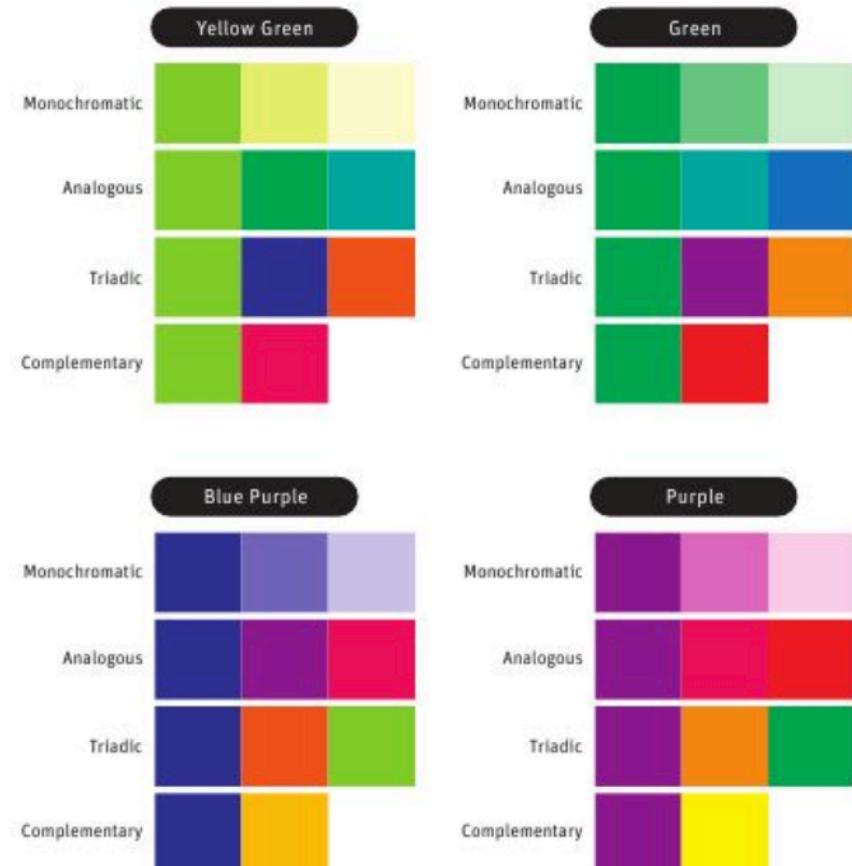
# Element 3 – Color

## Some guidelines of color in design

Once you choose a color –  
create color harmony

Suggestions for building a  
color palette –

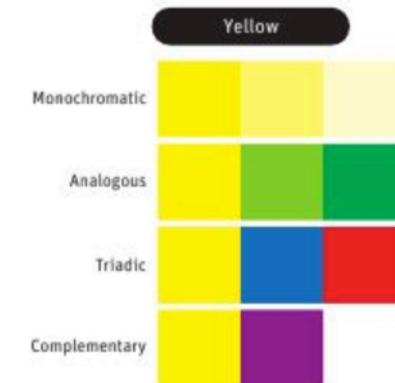
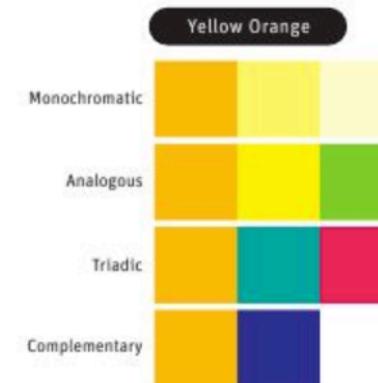
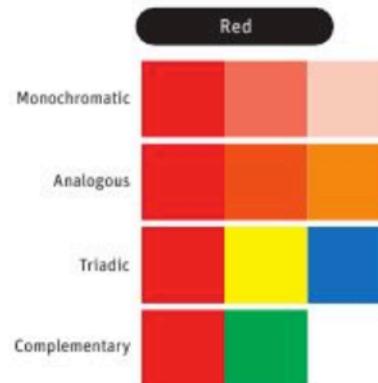
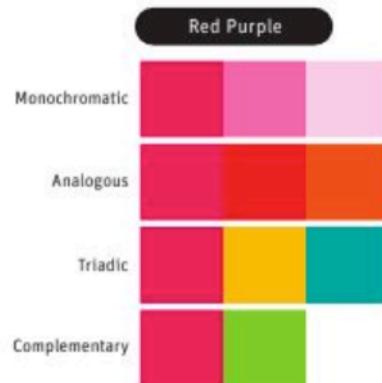
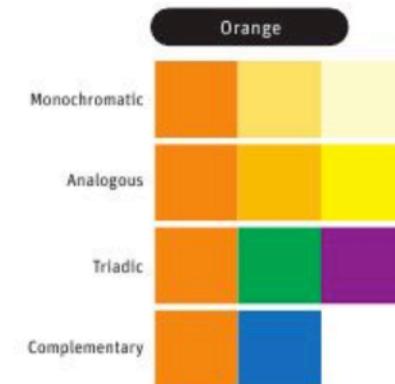
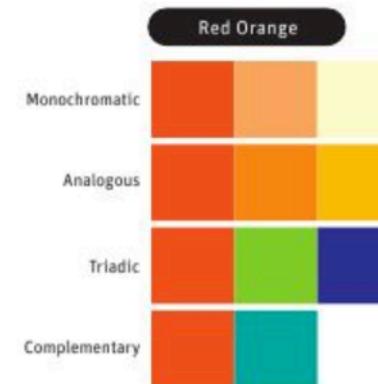
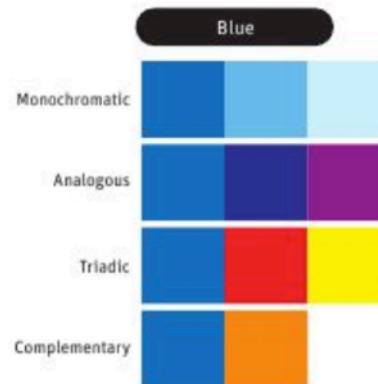
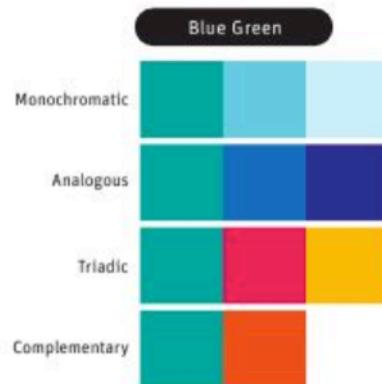
- Choose a primary color – and then accent around it
- Do not overload on colors, usually 5 or less is good enough
- Play around – what looks good? What accentuates difference?



# Element 3 – Color

## Some guidelines of color in design

Example of color choice in action – these are classic pallets to put together – you can find more of course online



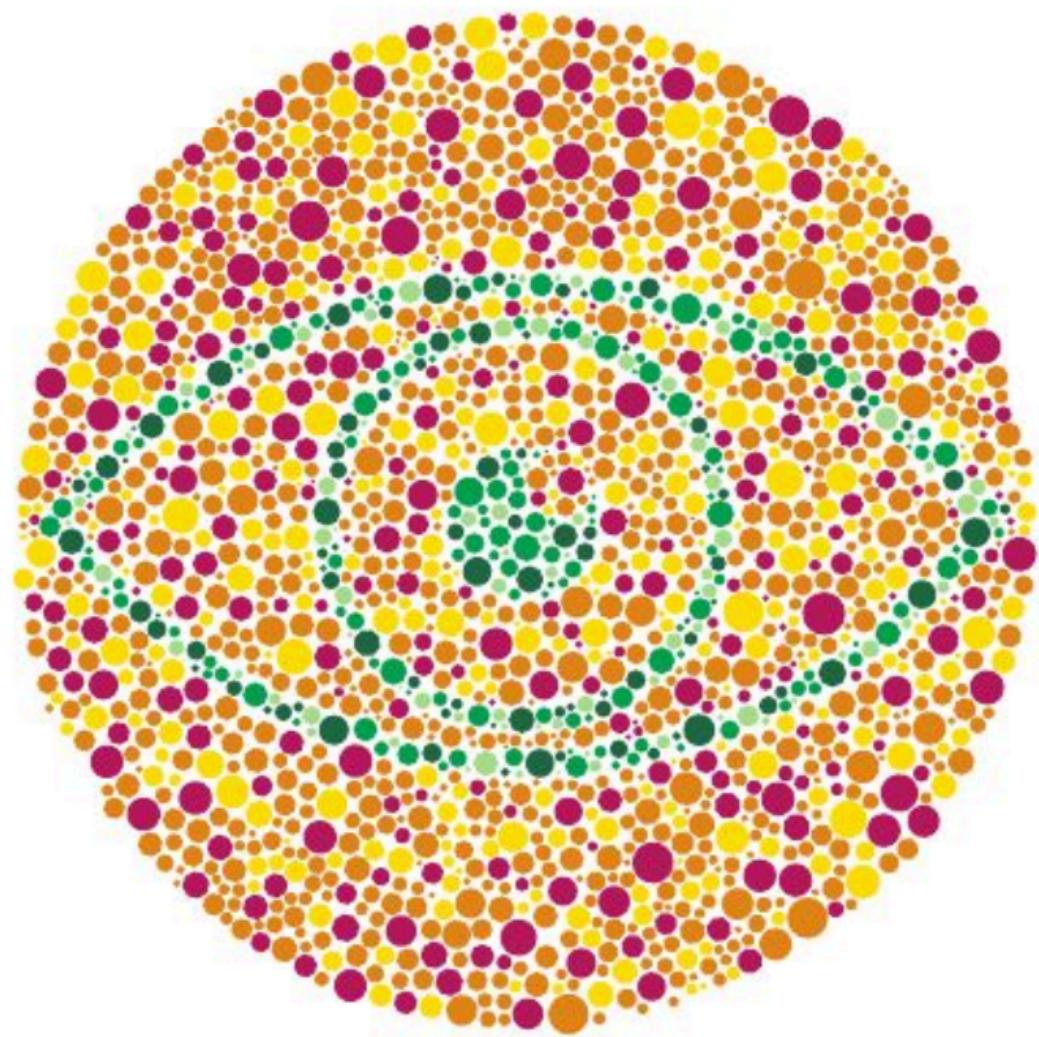
# Element 3 – Color

Some guidelines on color in design

People see colors differently

Good colors to *avoid*:  
Red & Green in combo

Resource: You can get  
“hex codes” and palettes  
from a variety of websites  
– my favorites [\[1\]](#), [\[2\]](#), [\[3\]](#)



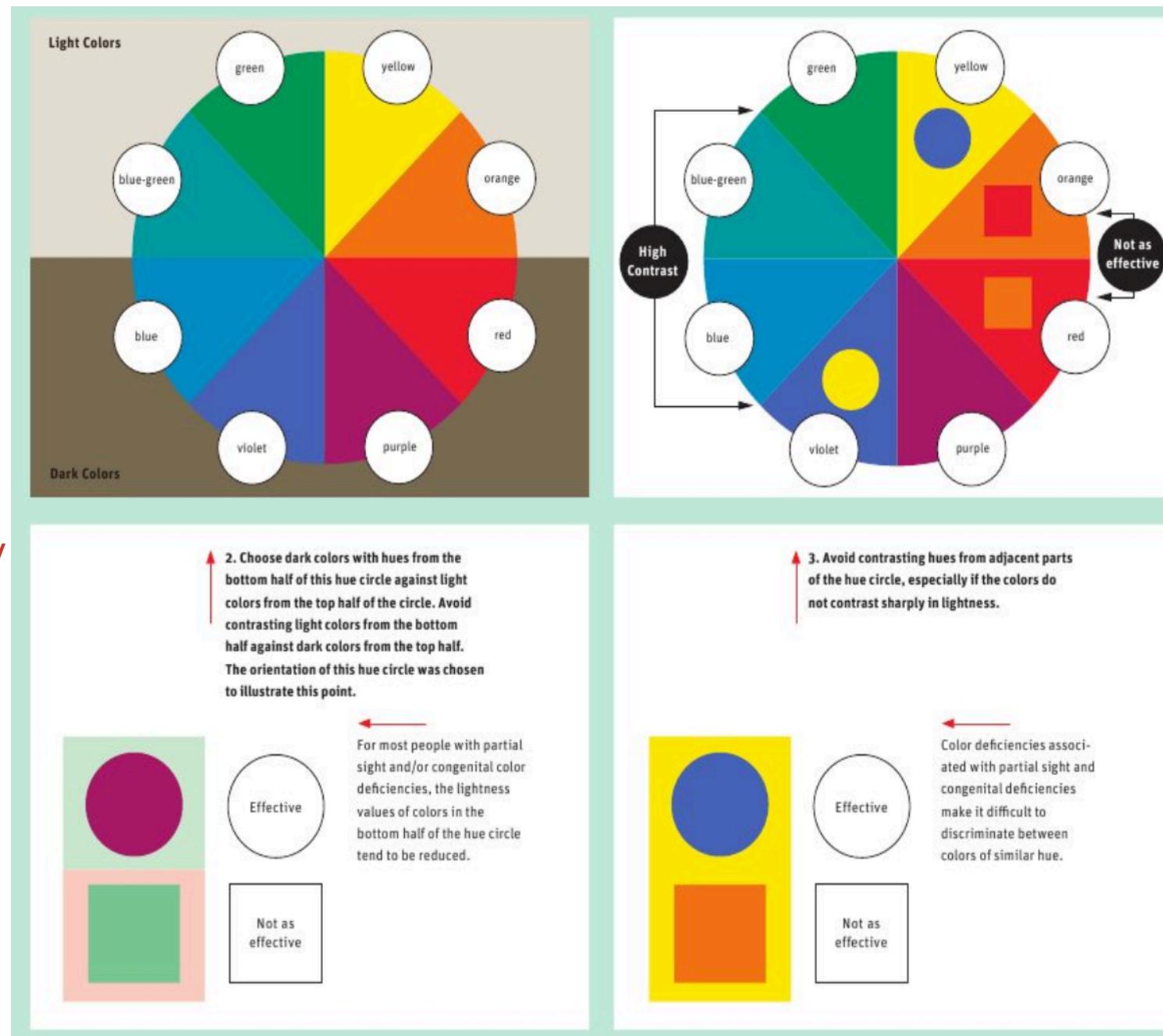
# Element Color

## Some guidelines of color in design

People see colors differently

Try to vary in hue AND saturation

Feel free to experiment!

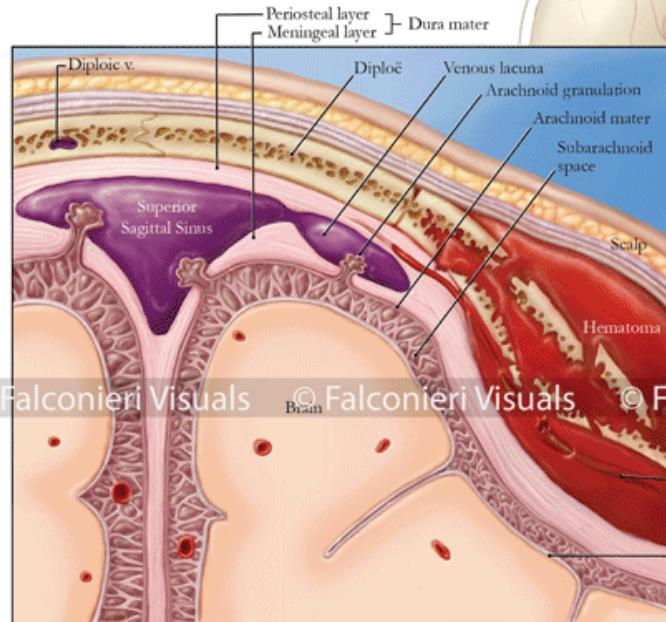


# Element 4 – Layering

## Add layers to express meaning

### Role of the Middle Meningeal Artery in Extradural Hematoma

The middle meningeal artery is often the source of blood during extradural hematoma. Due to its location adjacent to the bone in more lateral regions of the skull, the middle meningeal artery is easily damaged by head trauma. Blood accumulation between dura mater and bone creates pressure that may result in brain injury or death.



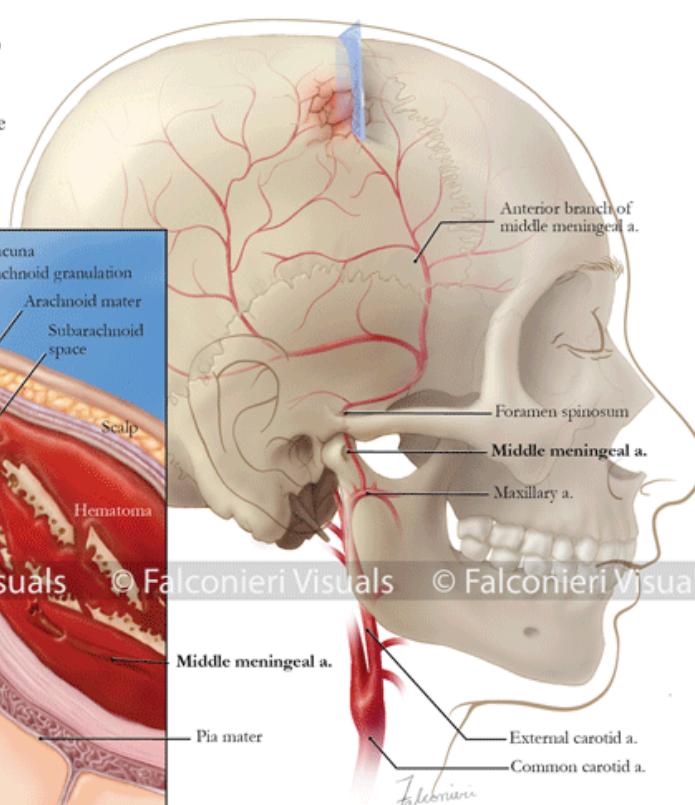
© Falconieri Visuals

© Falconieri Visuals

© Falconieri Visuals

© Falconieri Visuals: Contact Veronica@FalconieriVisuals.com for usage licenses and inquiries.

Falconieri\_meningeal\_anatomy



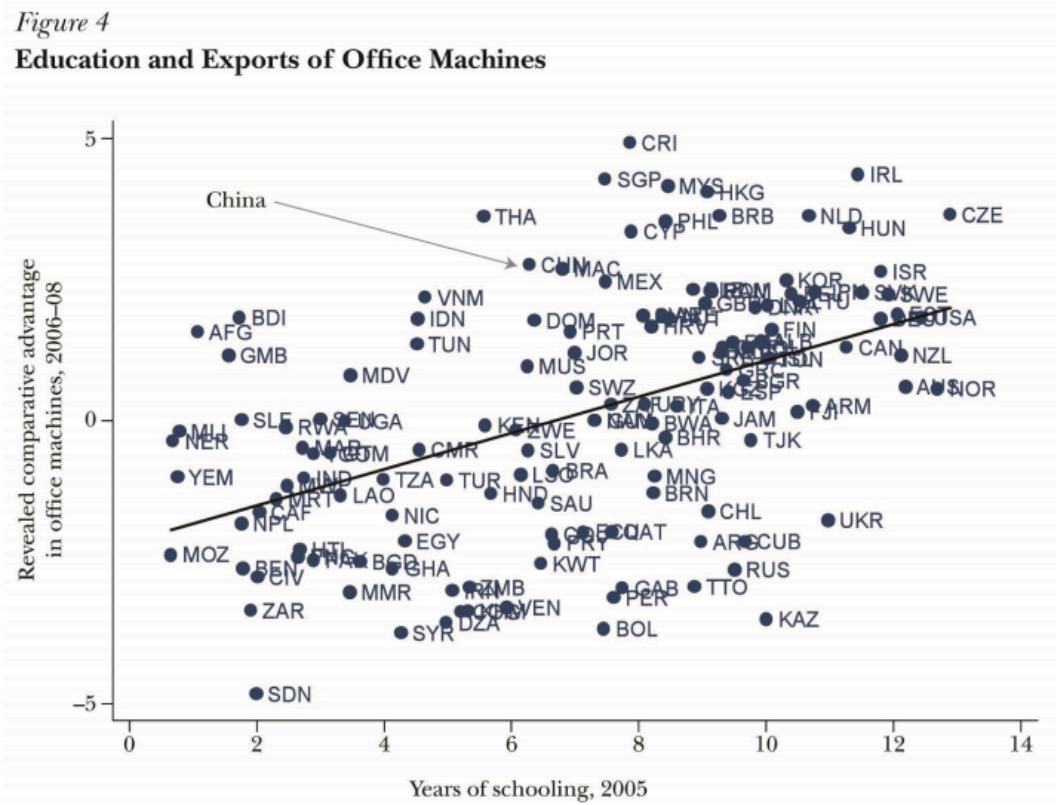
From V.  
Falconieri –  
biological  
illustrator

[http://  
falconierivisuals.  
com/](http://falconierivisuals.com/)

Based on Frankel and DePace, 2012.

# Element 5 – Refine

Once you have a good figure – simplify to express meaning  
**Reduce Clutter**



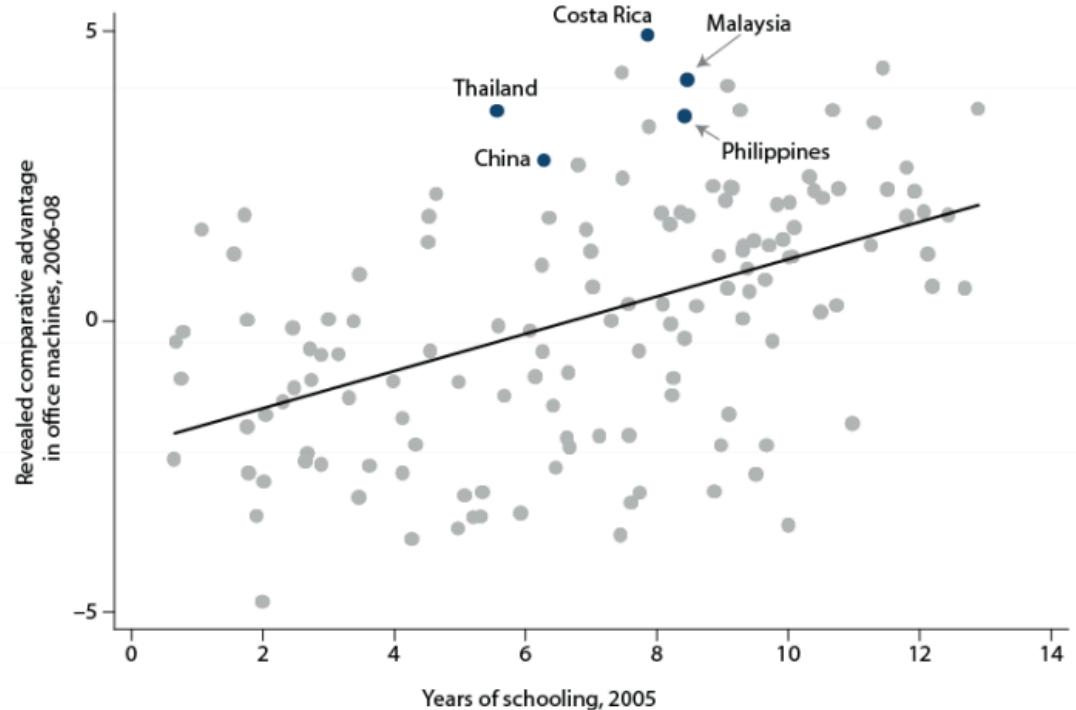
Based on Frankel and DePace, 2012.

Adapted from Telling Visual Stories About Data, Congressional Budget Office, Fontaine, 2014.

# Element 5 – Refine

Once you have a good figure – simplify to express meaning  
**Reduce Clutter**

*Figure 4*  
Education and Exports of Office Machines



Based on Frankel and DePace, 2012.

Adapted from Telling Visual Stories About Data, Congressional Budget Office, Fontaine, 2014.

# What we've learned

When starting a visualization for data consider what is the main point of my graphic and what is my audience?

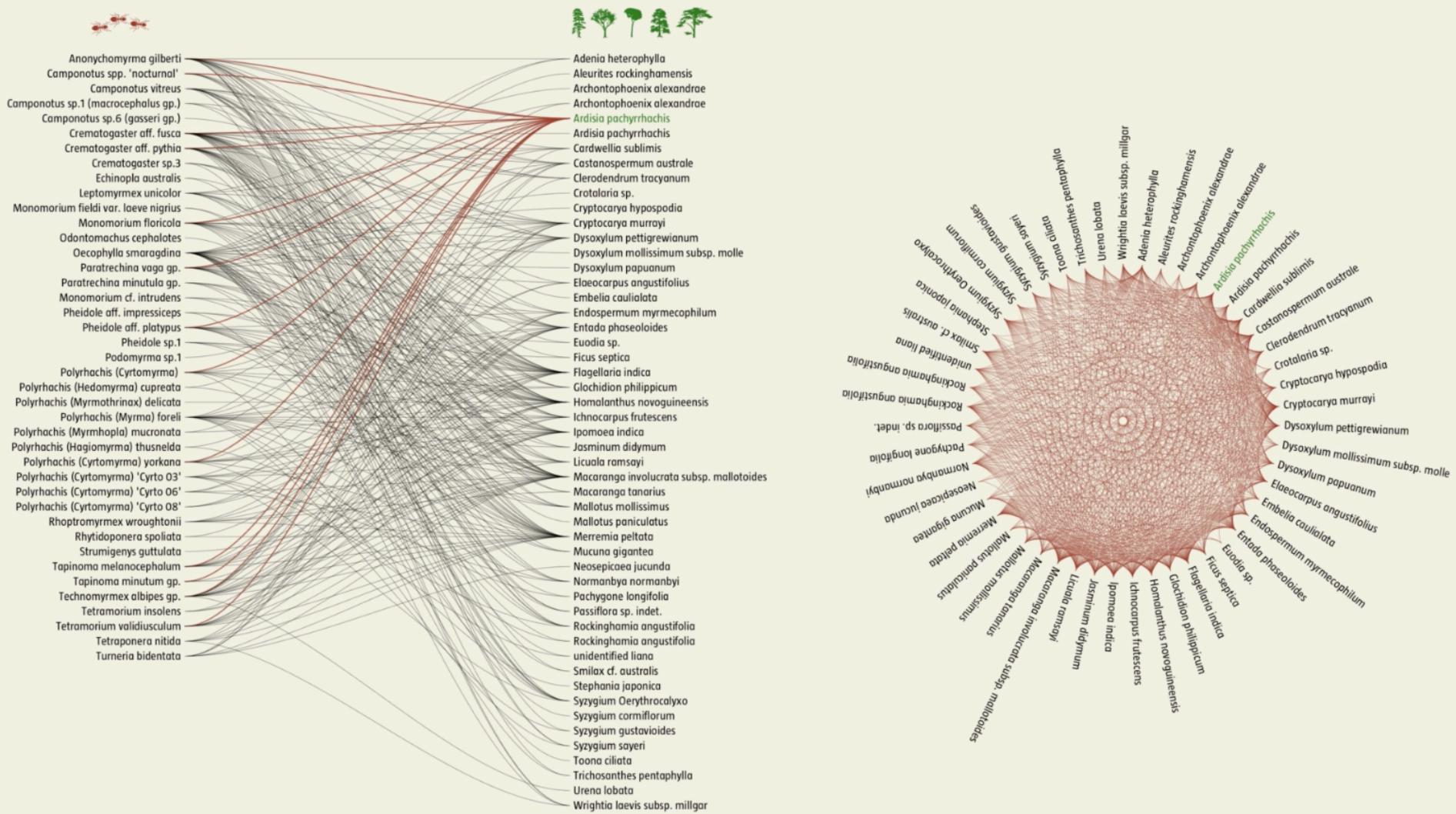
- Is my graphic **explanatory** or **exploratory**?
- How will it be **used**?
- What is the first thing you want your audience to see?

When making a figure consider:

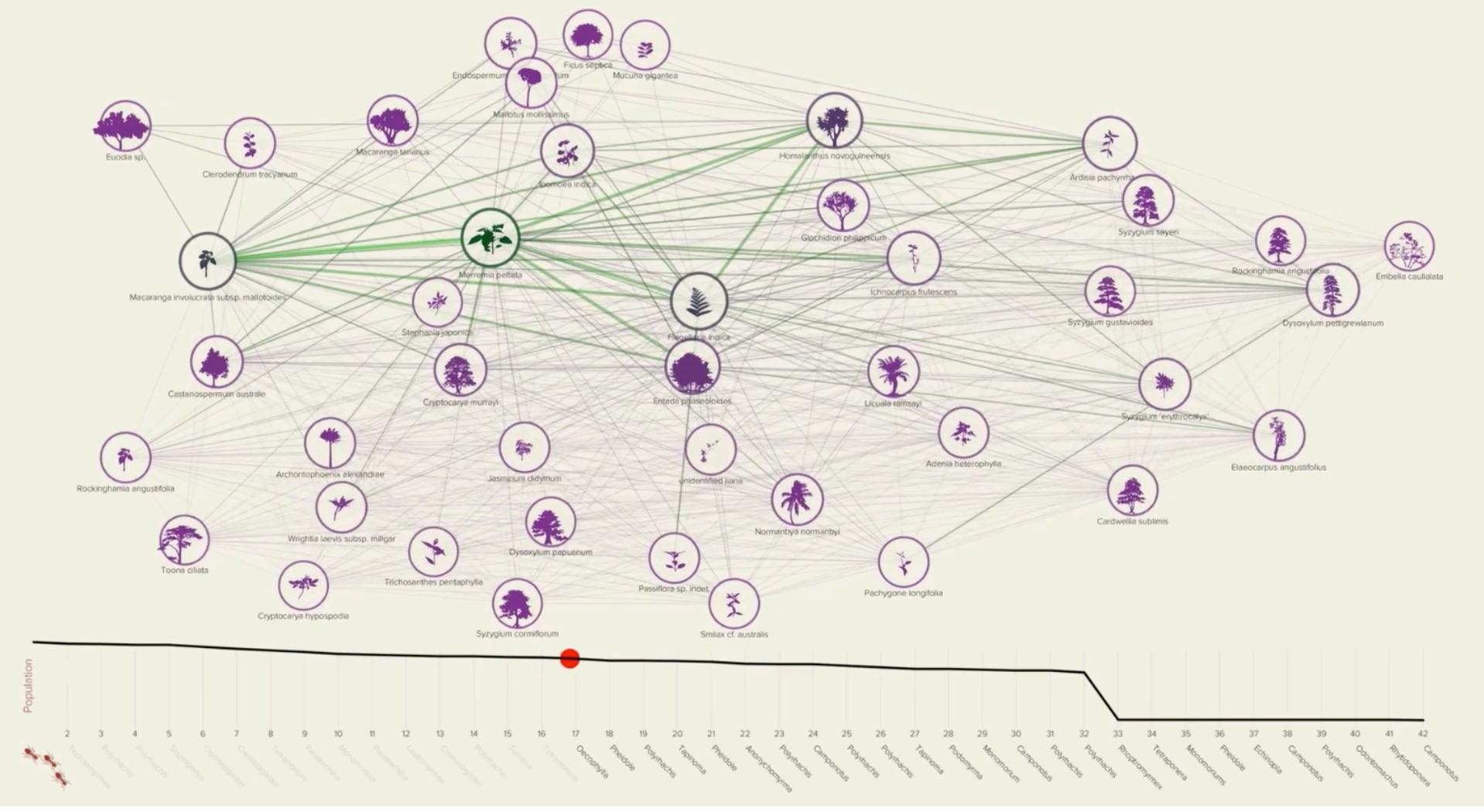
- Composition
- Abstraction
- Coloring (make sure to consider colorblind palettes!)
- Layering
- Refining

# Let's put what we've learned into practice

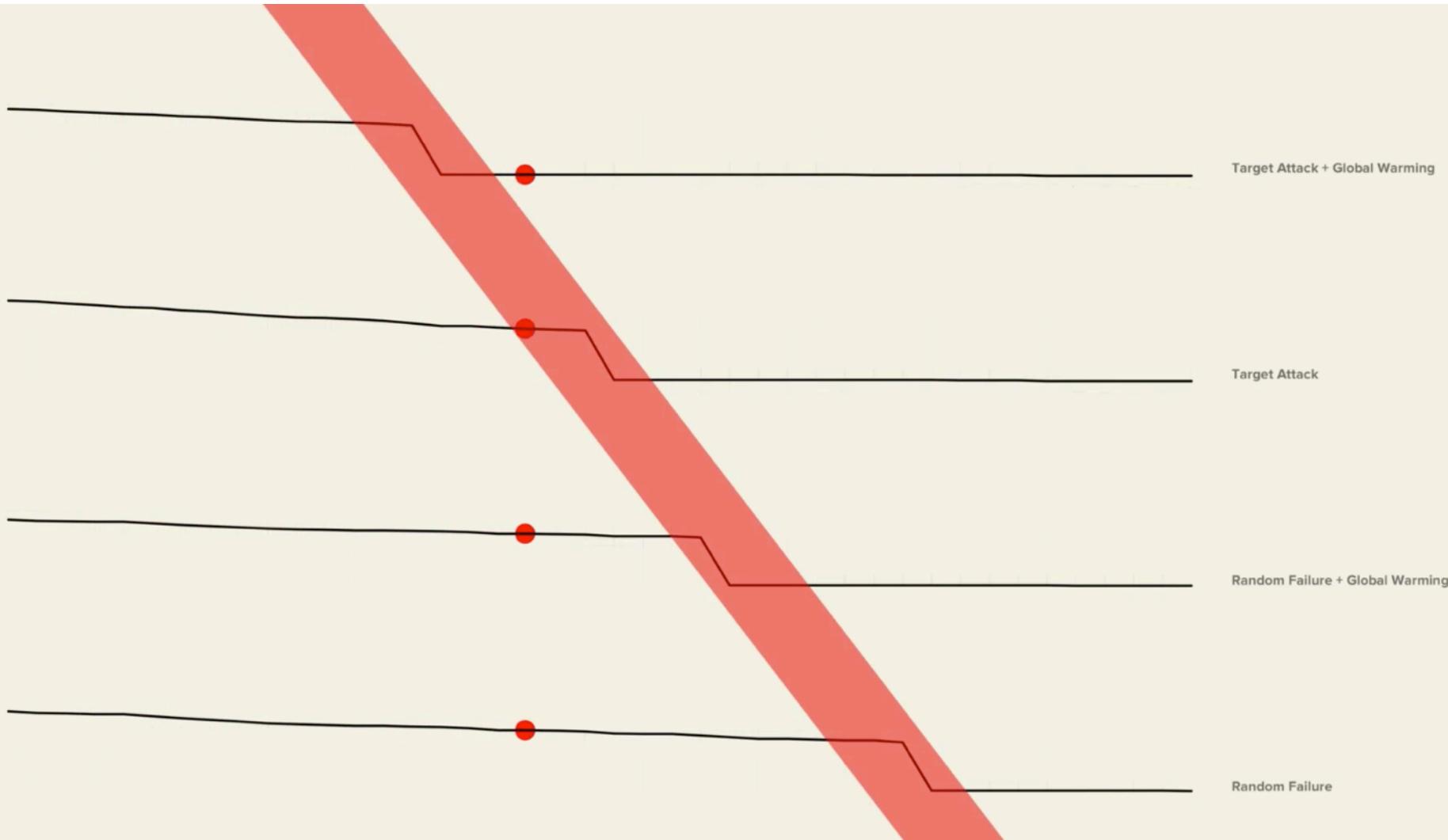
- Consider these as we watch the following video -  
[http://players.brightcove.net/679256133001/NkgrDczuol\\_default/index.html?videoid=5373954480001](http://players.brightcove.net/679256133001/NkgrDczuol_default/index.html?videoid=5373954480001)
- This won the NSF Expert's Choice award in 2017 for best video visualization.
- Think as we watch this how these use the elements to address their audience
  - Composition
  - Abstraction
  - Coloring
  - Layering
  - Refining



Source: Network Earth, Mauro Martino, Jianxi Gao, Baruch Barzel, Albert-László Barabási. Narration: Shamini Bundell



Source: Network Earth, Mauro Martino, Jianxi Gao, Baruch Barzel, Albert-László Barabási. Narration: Shamini Bundell



Source: Network Earth, Mauro Martino, Jianxi Gao, Baruch Barzel,  
Albert-László Barabási. Narration: Shamina Bundell

# In Python -

We will be using the basics plotting / visualization package called [matplotlib](#) (the documentation is amazing)

- This is commonly used in Python as:
  - import matplotlib.pyplot as plt
- This is not the *ONLY* plotting package – the other popular one for data science and stats is seaborn
- In the following sessions we will see a *brief* teaser of what matplotlib can do – the Jupyter notebook includes hyperlinks to resources and other information as you get started
- Also feel free to check out the [Github here](#)
  - this is lab section of a new class in Climate and Space (4 credits for undergrad and grad) that teaches data visualization and statistics – novices are welcome!