Lab 3a - Good and Bad Visualizations

YourNameGoesHere

For class Tuesday, February 15th

Good and Bad Visualizations

If you Google `examples of good visualizations` or `bad visualizations` or any combination, you will find a plethora of examples. Understanding the taxonomy can help in identifying poor visualizations and what aspects are poor. It's even better when you can think of appropriate ways to present the material yourself. It is often possible to mention an area of improvement, even if the basic structure of a graphic is sound.

In this lab, we'll be considering some visualizations to get a sense of what is out there and what to be careful of. As you explore these examples, you'll look for properties you like and issues you see. You'll also be deciding if you think each visualization is good or bad, and you may not agree with your groupmates/peers.

Note: It is possible to make a bad visualization with ggplot2 (too much information, unreadable, etc.), but it is harder than when free-drawing. Using appropriate software makes it much harder to violate the area principle, for example.

You'll be critiquing the figures shown at the end of the lab with your groups. First, let's review the Git workflow.

Git Workflow Review

1. Before editing this file, verify you are working on the copy saved in *your* repo for the course (check the filepath and the project name in the top right corner).

2. Before editing this file, make an initial commit of the file to your repo to add your copy of the lab.

3. Change your name at the top of the file and get started!

4. You should *save, knit, and commit* the .Rmd file each time you've finished a question, if not more often.

5. You should occasionally *push* the updated version of the .Rmd file back onto GitHub. When you are ready to push, you can click on the Git pane and then click **Push**. You can also do this after each commit in RStudio by clicking **Push** in the top right of the *Commit* pop-up window.

6. When you think you are done with the lab, save the pdf as "*YourFirstInitialYourLast-Name_thisfilename*.pdf" before committing and pushing. For example, I would save this file as AWagaman_Lab3a.pdf.

Four figures are included at the end of this lab. For each figure, examine the figure individually and then discuss the questions as a group once everyone has had a chance to look it over. You may want to look at the compiled .pdf or the .png files to enlarge the figures compared to how they appear in RStudio.

Figure 1 - Genetics

1. What story does the figure tell?

2. As best you can, identify the variables included in the figure, and identify how they're represented (e.g. visual cue).

3. What, if anything, do you like about the original figure?

4. What, if anything, were some key problems with the original figure?

5. Brainstorm another method of displaying the information in this graph. Do your best to describe what you would do in an adaptation of the visualization. For example, what would you label your axes and which visual cues would you use? You may suggest breaking the graphic into 2 different plots (still combined to tell the story) if necessary.

6. In what way(s) is your suggested plot an improvement on the original visualization?

7. In what way(s), if any, is the main message of the plot still unclear/hidden even in your adaptation?

8. What is the group consensus about the original plot? Was it a good or bad visualization? (Note, you can brainstorm ways to improve both the good and bad ones!)

Figure 2 - Napoleon's March

1. What story does the figure tell?

2. As best you can, identify the variables included in the figure, and identify how they're represented (e.g. visual cue).

3. What, if anything, do you like about the original figure?

4. What, if anything, were some key problems with the original figure?

5. Brainstorm another method of displaying the information in this graph. Do your best to describe what you would do in an adaptation of the visualization. For example, what would you label your axes and which visual cues would you use? You may suggest breaking the graphic into 2 different plots (still combined).

6. In what way(s) is your suggested plot an improvement on the original visualization?

7. In what way(s), if any, is the main message of the plot still unclear/hidden even in your adaptation?

8. What is the group consensus about the original plot? Was it a good or bad visualization? (Note, you can brainstorm ways to improve both the good and bad ones!)

Figure 3 - AAUP Employment

1. What story does the figure tell?

2. As best you can, identify the variables included in the figure, and identify how they're represented (e.g. visual cue).

3. What, if anything, do you like about the original figure?

4. What, if anything, were some key problems with the original figure?

5. Brainstorm another method of displaying the information in this graph. Do your best to describe what you would do in an adaptation of the visualization. For example, what would you label your axes and which visual cues would you use? You may suggest breaking the graphic into 2 different plots (still combined).

6. In what way(s) is your suggested plot an improvement on the original visualization?

7. In what way(s), if any, is the main message of the plot still unclear/hidden even in your adaptation?

8. What is the group consensus about the original plot? Was it a good or bad visualization? (Note, you can brainstorm ways to improve both the good and bad ones!)

Figure 4 - British Army Mortality

1. What story does the figure tell?

2. As best you can, identify the variables included in the figure, and identify how they're represented (e.g. visual cue).

3. What, if anything, do you like about the original figure?

4. What, if anything, were some key problems with the original figure?

5. Brainstorm another method of displaying the information in this graph. Do your best to describe what you would do in an adaptation of the visualization. For example, what would you label your axes and which visual cues would you use? You may suggest breaking the graphic into 2 different plots (still combined).

6. In what way(s) is your suggested plot an improvement on the original visualization?

7. In what way(s), if any, is the main message of the plot still unclear/hidden even in your adaptation?

8. What is the group consensus about the original plot? Was it a good or bad visualization? (Note, you can brainstorm ways to improve both the good and bad ones!)
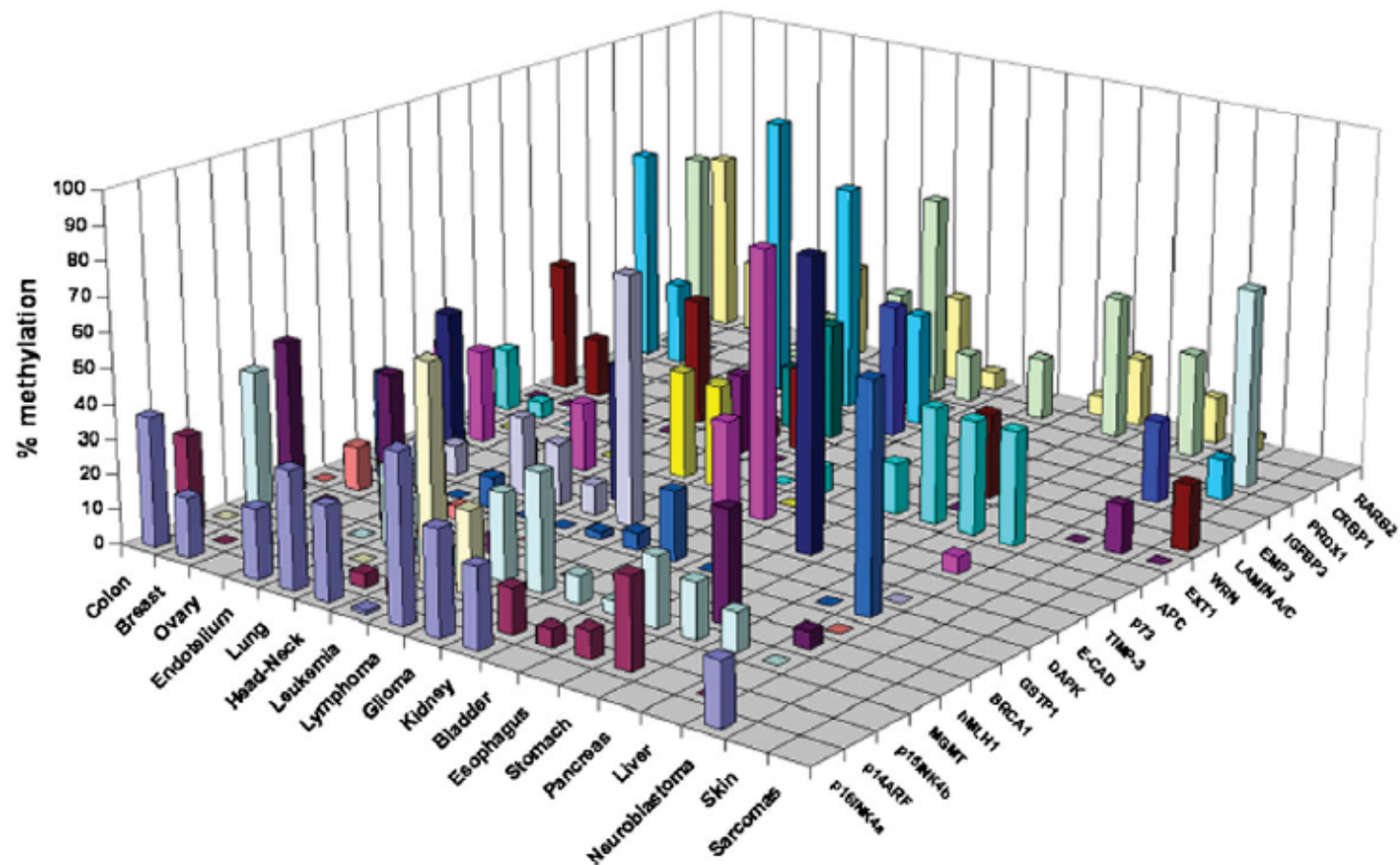
Figure 1. A CpG island hypermethylation profile of human cancer. *Y*-axis, frequency of hypermethylation for each gene in each primary.
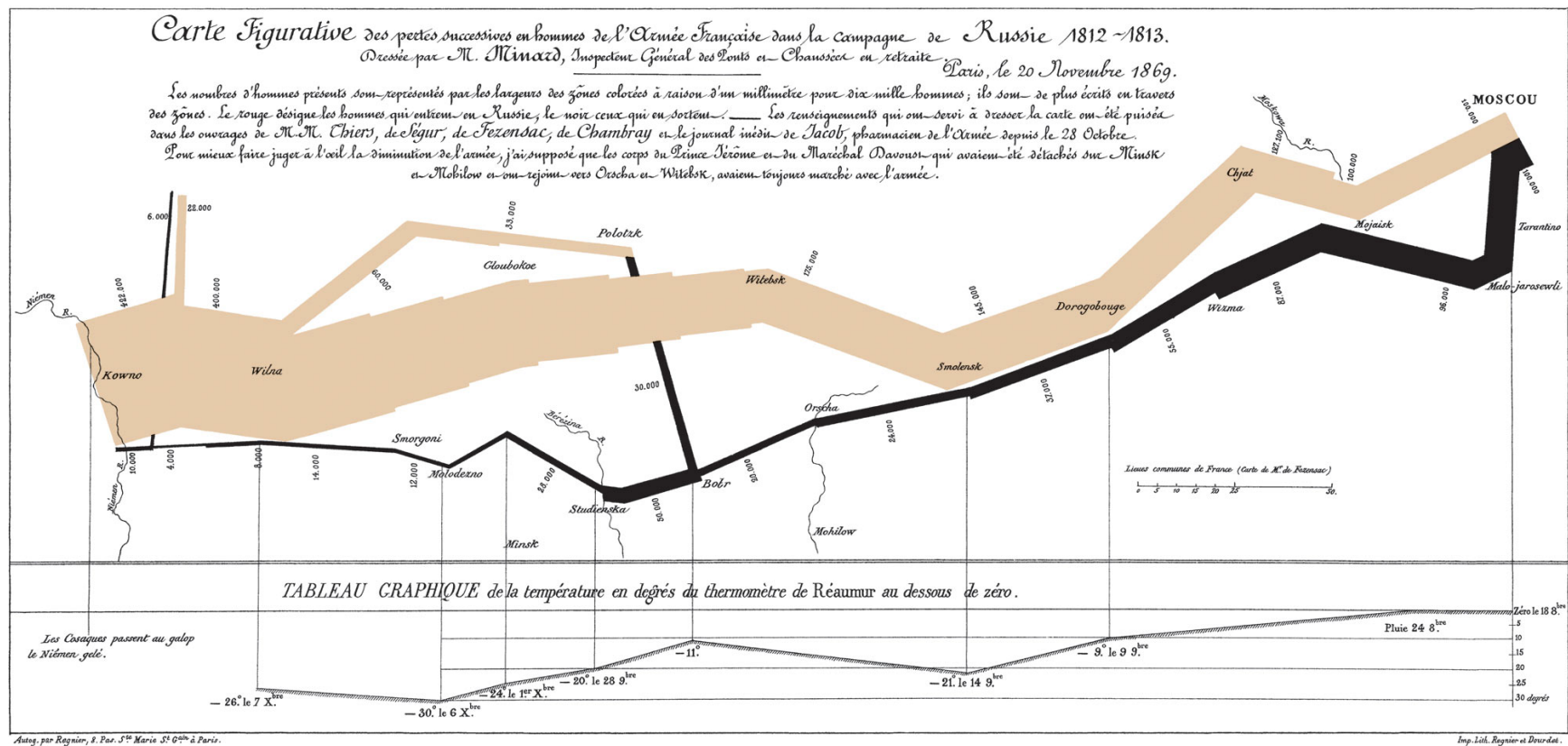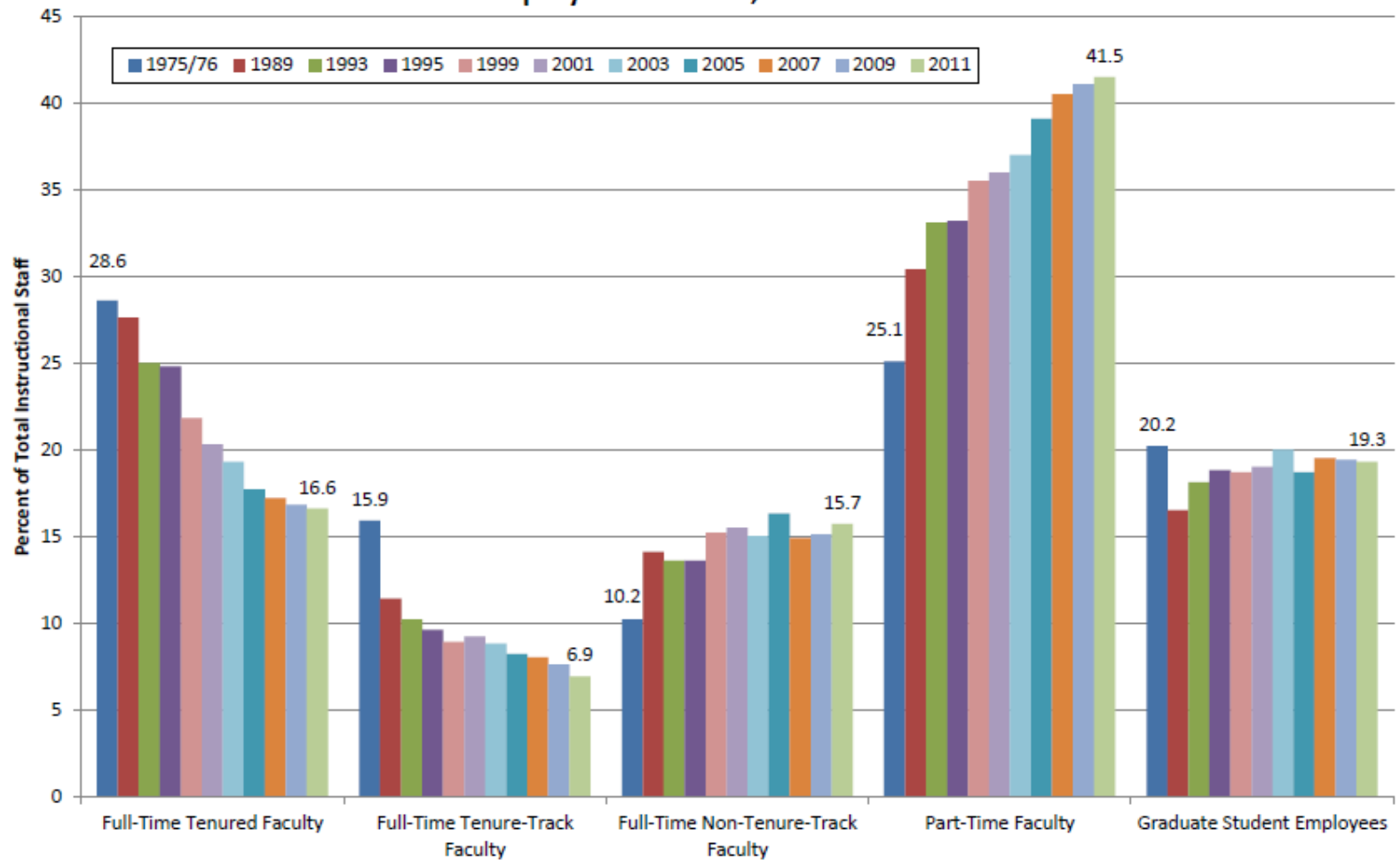
Figure 1: Human Molecular Genetics Results Figure.

Figure 2: Minard's Plot of Napoleon's March.

Figure 1

Trends in Instructional Staff Employment Status, 1975 and 1976 to 2011

Notes: Figures in this chart have been updated from those published by AAUP in 2013. 1975-76: Figures for full-time faculty are for 1975 and are estimated; all other figures are for 1976. For 1976, full-time and part-time graduate student employees are included; in later years all were part-time by definition.

Source: US Department of Education, IPEDS Fall Staff Survey. Tabulation by John W. Curtis, American Association of University Professors, Washington, DC.

Figure 3: AAUP Staff-Employment.

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST

2.
APRIL 1855 TO MARCH 1856

1.
APRIL 1854 TO MARCH 1855

JUNE
MAY
JULY
APRIL
AUGUST
MARCH
SEPTEMBER OCTOBER
FEBRUARY
1854
JANUARY DECEMBER NOVEMBER

AUGUST
JULY
JUNE
MAY
BULGARIA
APRIL 1854
SEPTEMBER
CRIMEA
OCTOBER
MARCH 1855
NOVEMBER
FEBRUARY 1855
DECEMBER
JANUARY 1855

The Areas of the blue, red, & black wedges are each measured from
   the centre as the common vertex
The blue wedges measured from the centre of the circle represent area
   for area the deaths from Preventible or Mitigable Zymotic Diseases, the
   red wedges measured from the centre the deaths from wounds, & the
   black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov.ʳ 1854 marks the boundary
   of the deaths from all other causes during the month
In October 1854, & April 1855, the black area coincides with the red,
   in January & February 1856, the blue coincides with the black
The entire areas may be compared by following the blue, the red & the
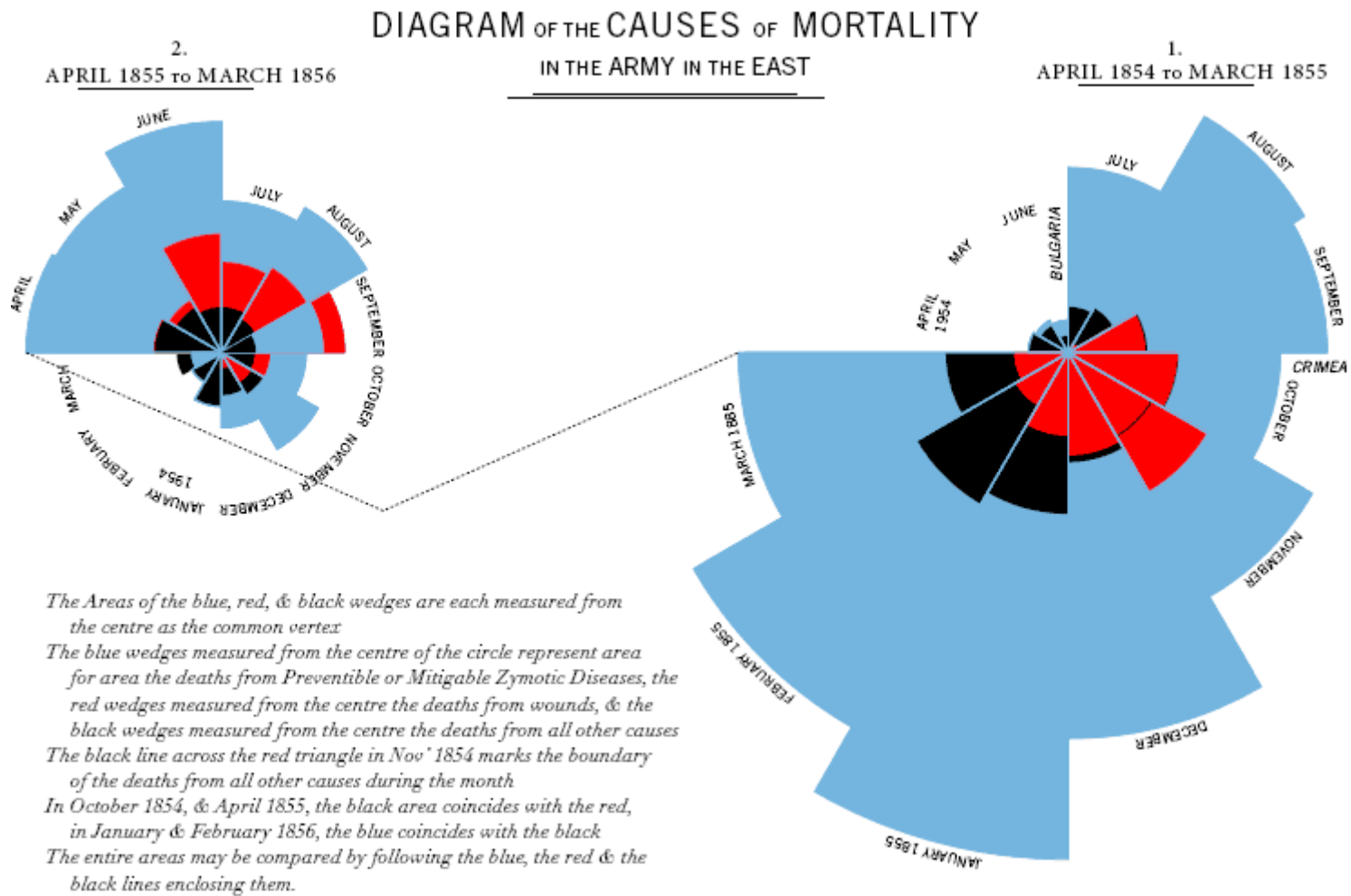   black lines enclosing them.

Figure 4: Florence Nightingale (1820-1910) is famous not only for her push for more sanitary conditions in hospitals and homes, but also for her work in statistics. She created this now famous "coxcomb" graph to visualize seasonal sources of patient mortality in a British field hospital during the Crimean War. A Sanitary Commission was created in March of 1855, which is the reason for the split in the two graphs after that month.

References for Figures

1. Manel Esteller, Epigenetic gene silencing in cancer: the DNA hypermethylome, Human Molecular Genetics, Volume 16, Issue R1, 15 April 2007, Pages R50–R59, https://doi.org/10.1093/hmg/ddm018
2. Charles Minard's map of Napoleon's march. https://en.wikipedia.org/wiki/Charles_Joseph_Minard
3. Curtis, John W. "The employment status of instructional staff members in higher education, Fall 2011." American Association of University Professors (2014): 2011-12.
4. Florence Nightingale's coxcomb graph. https://en.wikipedia.org/wiki/Florence_Nightingale

Unused but potentially interesting for you

1. Nathan Yau's blog - https://flowingdata.com/
2. Friendly, Michael (2001), Gallery of Data Visualization, Electronic document, http://www.datavis.ca/gallery/
3. Examples of good and bad visuals compiled by CS251 students at Colby College (Prof. Stephanie Taylor) https://cs.colby.edu/courses/S14/cs251/goodbad.php