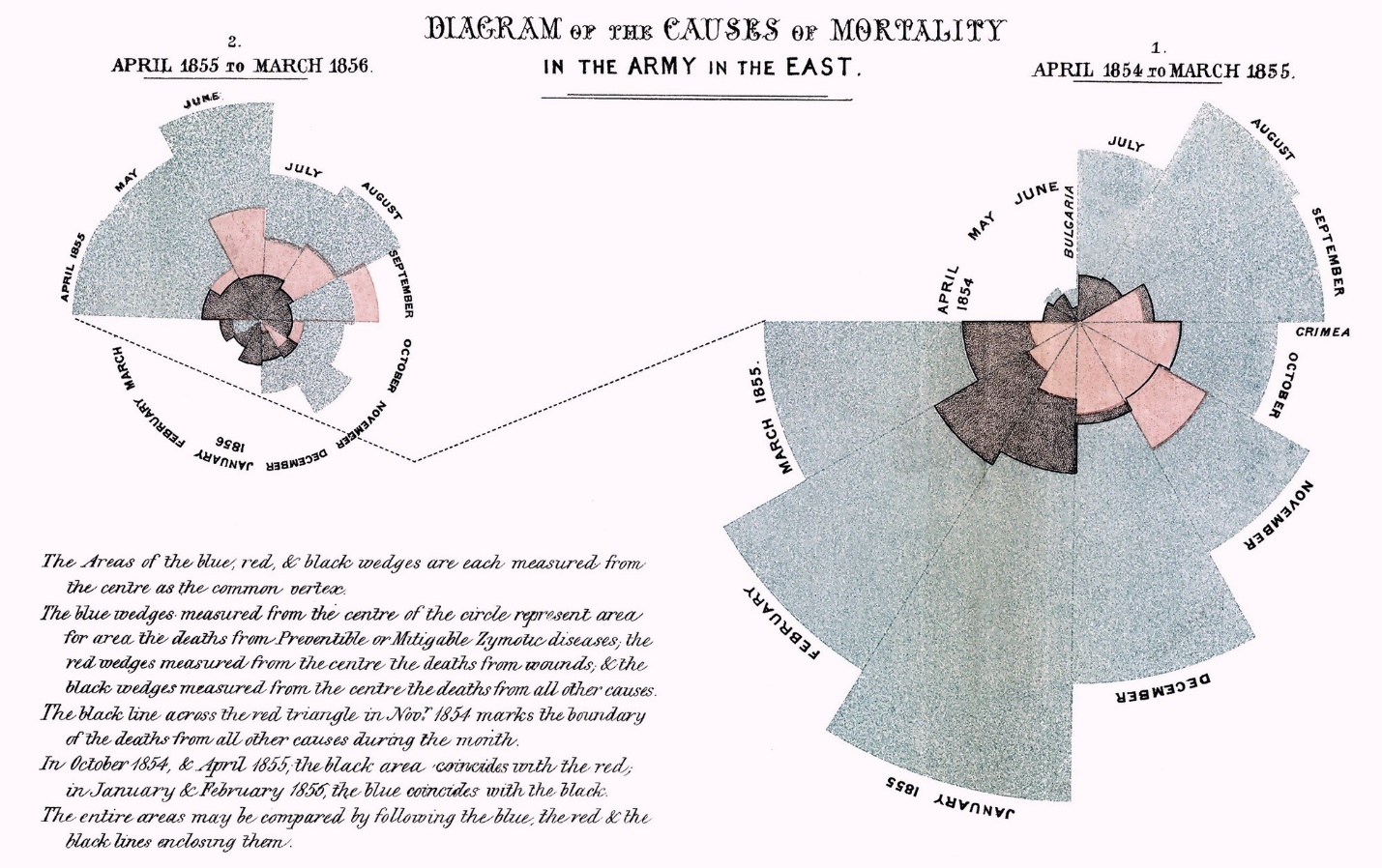
1. Tidying data is making your data pretty, consistent and easy to work with. It involves ensuring each column is a variable and each row is an observation. This should make the data easy to work with and transform.
2. Transforming data is narrowing in on data you care about (by filtering or sorting or a similar operation), creating new variables which are a function of one or multiple other variables, and calculating statistics of interest such as mean, counts, standard error or others.
3. VIS is both very important but also a skill you will acquire more and more as time goes on. VIS is especially important for conveying information to others. Whether this be giving information to supervisors, clients or peers it is important to make this data accessible and since humans are best at learning visually and good at seeing patterns in front of them, VIS is an important means for this end. For example, one career I was interested in was sports analytics. In the world of sports analytics, it is important to convey information to those who need it. For example, it is good for conveying average homeruns over time, where batters are likely to hit the ball, or almost any other statistic regarding sports.
4. 

Data is from <https://www.tableau.com/learn/articles/best-beautiful-data-visualization-examples>

I really like this data and the way it is visualized. The way that the data works is red data is death from wounds, black data is deaths from battle, and blue is deaths from poor hospital practices. It is also broken down by month between April 1854 and March 1855 in the Crimean War of the 1850s. What it shows is that battles and wounds caused some deaths, but the vast majority of deaths were caused by errors by hospital malpractice.

1. Adding –global will set that as you commit email for all directories, rather than just the current directory.
2. Filter, which filters the data, arrange, which sorts the data, select, which allows you to choose certain variables to choose or remove, mutate, which allows you to create new columns and summarize, which allows you to collapse data into information like the mean
3. Cumulative provides the calculation of all the data, the rolling aggregate is just over a specified number of data points/time
4. Provides the rank of items in their data set (e.g. if you were to sort them, where would that specific item be in the list)
5. Come up with an idea, come up with a plan to develop the idea, implement (code) the idea, test and fix issues and adapt as problems arise, then give the product to the consumer.

For example, the idea could be to create a visualization of covid cases across the US, the plan would be to use R and shiny to create this visualization, they would then work together (potentially on different parts) of the product, they would then test and fix issues that arise then they would present the data to the person who wanted this project.

1. A generic software product development team would want their product to be easy to use by a wide variety of people/companies who have an issue related to what their product is designed to solve. It will need to be adaptable to fit a variety of consumer needs.

A custom product development team wants their product to work for only their specific client. It needs to work exactly to their clients specifications and work to fix their needs.

In practice this means that generic software users will not have everything work for them and they will not get to dictate how the software is designed, instead they will be required to adapt and create their own tools in order for the software to cater to their needs better.