The Data Life Cycle

generation	collection	processing	storage	management	analysis	visualization	interpretation
"People generate data: every search query we perform, link we click, movie we watch, book we read, picture we take, message we send, and place we go contribute to the massive digital footprint we each generate" *[Think also of historical source documents]*	*"Not all data generated is collected, perhaps out of choice because we do not need or want to, or for practical reasons…. Deciding what to collect defines a filter on the data we generate"*	*"everything from data cleaning, data wrangling, and data formatting to data compression, for efficient storage, and data encryption, for secure storage"*	*"the bits are laid down in memory"*	*"We are careful to store our data in ways both to optimize expected access patterns and to provide as much generality as possible. We need to create and use different kinds of meta data for these dimensions of heterogeneity to maximize our ability to access and modify the data for subsequent analysis"*	*"all the computational and statistical techniques for analyzing data for some purpose: the algorithms and methods that underlie artificial intelligence (AI), data mining, machine learning, and statistical inference, be they to gain knowledge or insights, build classifiers and predictors, or infer causality"*	*"helps present results in a clear and simple way that a human can readily understand and visualize"*	*"we provide the human reader an explanation of what the picture means. We tell a story explaining the picture's context, point, implications, and possible ramifications"*

Jeannette M. Wing (2019) "The Data Life Cycle" *Harvard Data Science Review*

Write down 3 big challenges you've faced with data (or lessons learned). Add them below.

generation	collection	processing	storage	management	analysis	visualization)	interpretation)
use/deployment				throughout			