Computational Data Practices Planning Guide

# Essential Question:

# Creating: Generating data from tools or observation

## Decomposition

1. *Break it down*: What are the key types/sources of data that we might use to answer the essential question? How will we generate this data? What metrics can we use?
2. *Evaluate and choose*: Is the data primary or secondary? Which of the data sources will best help to answer our essential question?

## Algorithm Building

1. *Generalize*: Which process will help us to effectively generate data to solve this problem? What are the rules or processes for creating data?
2. *Identify Conditions*: What conditions (e.g., if this, then that) need to be implemented to gather the data?
3. *Put it in order*: What is the most efficient sequence to follow to generate this data? Are there steps/processes that should run simultaneously? Write out the steps below.

***Automation***

1. *Identify tools*: What tools can we use/create to automate the generation of this data? What commands/functions in these tools will best help me automate data generation?

# Collecting: Gathering and recording data

## Pattern Recognition

1. *Identify patterns*: What are the intervals of the collected data? When does a data cycle end and when does it begin? When/How often do the data repeat?

***Abstraction***

1. *Identify essential elements*: How frequently do we need to collect data? What data, if removed, would prevent us from answering the essential question?
2. *Identify noise*: Is there certain information that is less helpful when trying to figure out how best to gather data?
3. *Create a model*: Make an outline or drawing of how to collect the essential elements.

***Algorithm Building***:

1. *Generalize*: What are the rules for collecting the data?
2. *Identify conditions*: Under what conditions will you collect the data? Under what conditions will you NOT collect data?
3. *Put it in order*: What will be the steps/process we follow to collect the data? What is the order of these steps? Are there processes that need to run simultaneously?

***Automation***:

1. *Identify tools*: What tools could we use to collect the data more efficiently? More accurately?

#

#

# Manipulating: Sorting, filtering, cleaning, normalizing, and combining data sets

***Decomposition***:

1. *Break it down*: What variables can we use to filter/sort the data? What are the different ways we might filter/sort the data? Does the data need to be scaled? Is there a transformation to apply to the data (e.g., log10, log2, logn, exp)?
2. *Evaluate and choose*: What sub-questions/goals can we answer by reorganizing the data?

***Pattern Recognition***:

1. *Identify patterns*: In what ways does the data repeat? Are there bins or buckets we can use to group the data?

***Abstraction***:

1. *Identify essential elements: Can you filter the data to only show the most important data?*
2. *Remove noise*: Are there outliers that we can remove? How can we reduce noise in our data?
3. *Create a model*: Can we simplify the data by sorting/filtering/cleaning it to more closely match our mental model?

***Algorithm Building***:

1. *Generalize*: What are the rules or processes for manipulating data?
2. *Identify conditions*: Under what circumstances do we use each form of data manipulation? Is there a generalizable rule for which manipulation to use?
3. *Put it in order*: Is there an order to the steps in manipulating the data? What rules determine that order?

***Automation***:

1. *Identify tools*: What tools can help us to filter, sort, or clean the data? Is there a way to more easily manipulate the data by automating all or some of the process?

#

# Visualizing: Communicating results with a representation such as a graph or chart

***Decomposition***:

1. *Break it down*: What are possible ways could we visualize this data?
2. *Evaluate and choose*: Which visualizations would help us to answer our essential question? What new perspectives do different visualizations provide?

***Pattern Recognition***:

1. *Identify patterns*: What patterns can we see in the graphs/charts? How do the visualizations change as we add more data or manipulate the data?

***Abstraction***:

1. *Remove noise*: Can we adjust the type or scale of our visualization to focus on what we want to analyze? What is the clearest way of visualizing the data?
2. *Create a model*: How does this visualization match or contradict our mental model? Is there a way to visualize the data in a way that matches our mental model? Does the visualization suggest any changes we need to make to our mental model?

***Automation***:

1. *Identify tools*: What tools can we use to create these visualizations? Can we automatically update the visualizations as we add more data or manipulate the data? What commands/functions of the tool will help to automate data visualization?

# Analyzing: Extracting meaning from a data set for the purpose of drawing conclusions

***Decomposition***:

1. *Break it down*: What possible factors could help answer our question? Are there factors we haven’t considered yet that we should?

***Pattern Recognition***:

1. *ID patterns*: Are there factors that affect others? What is that relationship like? Do the data look random? Are they regular? Are there increasing or decreasing patterns? What “shape” is the data?

***Abstraction***:

1. *Remove noise*: Is there noise affecting our data? What kind of effect is that noise having on our analysis?
2. *Create a model*: How can we model the relationships between variables? Which of these relationships affects the outcome of our essential question? Is there a variable that affects the outcome more than the others?

***Algorithm Building***:

1. *Generalize*: What are the rules or processes for analyzing data? What formulas can we use to help us better discover relationships? Is there a formula we can use to describe the relationship between variables?

***Automation***:

1. *ID tools*: What tools can we use to analyze and model the data? What commands/functions of the tool will help to automate the analysis?