In this week’s discussion, I’d like to talk about the Seattle weather dataset(<https://www.kaggle.com/rtatman/did-it-rain-in-seattle-19482017)> which I have explored in week three’s discussion before. The dataset describes Seattle’s weather data from 1948/1/1 to 2017/12/17, including the date of the observation, the amount of precipitation, the maximum temperature for that day, the minimum temperature for that day and Boolean value whether rain was observed on that day.

In “A Design Space of Visualization Tasks”, the paper proposes that an individual task can be represented as a 5-tuple(goal, means, characteristics, target, cardinality). For more complicated compound tasks, they can be described hierarchically and factorized into multiple individual tasks. Furthermore, these compound tasks and individual tasks can be strung together to form workflows. In a word, workflows, compound tasks and individual tasks are three different levels of task description.

For the Seattle weather dataset, a possible task could be explore the relationship between daily maximum temperature’s trend and daily precipitation’s trend. This task can be broken down into two simple individual tasks: what’s the trend of daily maximum temperature and what's the trend of daily precipitation. These task could be exploratory and the means will be search. In this case, the compound task could be represented in enumeration as follows:

{(exploratory, search, trend, attribute(maximum temperature), all), (exploratory, search, trend, attribute(precipitation), all)}. For the workflow level, the task can be represented as:

{(exploratory, search, trend, attribute(maximum temperature), all), (exploratory, search, trend, attribute(precipitation), all)} => (relationship, compare, trend, attribute(maximum temperature)|attribute(precipitation),all).

After specify the tasks workflow, we could base on the contents of Ch 3 in the textbook to consider the kinds of visualization tools. The line graph could be a good visualization tool to explore the trends of maximum temperature and precipitation. What’s more, for comparison, just inspecting a single target in detail is not sufficient, a suitable visualization tool is combining the two high-level targets(trends) together and make them aligned so that it would be easier to observe the patterns and relationships.

In “Scatterplots: Tasks, Data, and designs” , the paper considers how to embed task and data characteristics in scatterplot design process. In their framework, they conclude the abstracted analysis tasks into 12 topics and categorize them into three different levels: object-centric, browsing and aggregate-level. At the same time, the framework also proposes clustering of design choices when building the scatterplot: point encoding like color, size, symbols; point grouping like representation type, positional binning, polygon enclosure, shape abstraction; point position like subsampling, displacement; graph amenities like grid lines, axis ticks.

For the Seattle dataset, the task of exploring the relationship between daily maximum temperature trend and daily precipitation trend will lies in browsing category, the task can be concluded as explore data’s trends and aggregate-level, identify correlation.

Based on the framework, the design choices could also be narrowed down once the task categories are decided. For the Seattle dataset, different point encoding like color or size could be used to distinguish between maximum temperature and precipitation; some graph amenities should also be taken into consideration to make the visualization clearer, like adding the grid lines, axis ticks, legend, trend lines and so on.