Exercise: Translate Questions into Tasks

Learning objective: Identify elements of an analytical question that need to be made more concrete. Translate high level questions into specific tasks that can be executed on a data set through data transformation and visualization.

How do you know if you are doing it right? You are doing this exercise right if you can easily identify which elements of a question are too ambiguous and if you can transform it into a specific data transformations and visual representations.

This exercise is based on the idea of transforming questions into tasks presented by Danyel Fisher and Miriah Meyer in their book Making Data Visual (Chapter 2). The main idea is that a fundamental preliminary step for effective visualization is to come up with good questions and translate them into actionable data transformation and visualization. This an often and yet critical step in effective data visualization.

In the following we are going to present a data analysis scenario. The scenario is made of a data set and a set questions. Your goal is to transform these questions into tasks by identifying proxy values and developing proxy tasks (as described in the book chapter mentioned above).

Scenario: NYU Vehicle Collisions

In this scenario you are assigned a data set describing vehicle collisions in NYC. The data set contains data for the last 5 years and it contains the following attributes:

- **Date:** the date of the collision
- **Time:** the time of the collision
- **Latitude:** the latitude of the location
- **Longitude:** the longitude of the location
- **Fatalities:** how many people lost their life
- **Injuries:** how many people got injured
- **Main cause:** the main cause of the collision

Main cause can have one of the following values: *Traffic Control Disregarded, Driver Inattention/Distraction, Failure to Yield Right-of-Way, Alcohol Involvement, Physical Disability, Backing Unsafely, Unsafe Speed.*

Your goal is to investigate the following question:

> "*What are the most dangerous spots in NYC*"?

Now assume you have another data set that collects information about weather in NYC. The data set has the following structures:

- **Date:** date of the measurement
- **Time:** time of the measurement
- **Precipitation:** precipitation amount
- **Wind speed:** recorded wind speed
- **Temperature:** recorded temperature

Your next goal is to investigate the following question:

> "*Does the weather have an effect on collisions*"?

Your next goal is to investigate the following question:

> "*How can we improve the situation*"?

From questions to tasks

For each the the questions presented above generate one or more proxy tasks, that is, tasks that can be executed on the data. For each task identify proxy values, how they are related to the original question and how they are computed (unless the values can directly be found in the data set).