Ideas for a Mission Control Data Analysis proof of concept

Data preparation

For this proof of concept we'll use the "US Weather Events" dataset at https://www.kaggle.com/sobhanmoosavi/us-weather-events

In order to simulate "refreshing" the data, let's turn one dataset into three by slicing it by StartTime(UTC):

1. One that includes only data from Jan 2016 through April 2020
2. One that includes only data from Jan 2016 through August 2020
3. The original dataset (Jan 2016 through December 2020)

Save these three files off separately as CSV's. Don't clean the data as part of this preparation. That way, we can demo what happens when we get updated data, e.g., replacing the data through 04/2020 with that through 08/2020.

First transformation: Data analysis

These analyses should be executed seriatim by a single driver script that accepts as an argument a path to the input file (see above).

Note: cleaning the data is not contemplated here, but feel free to clean it up to eliminate outliers. So long as it's done consistently, the specific cleaning doesn't matter.

1. Compute latitude deciles and count the number of events of each type (snow, rain, etc.) nationwide per decile per month. Save these in a CSV file with 120 rows (one per decile per month) and one column per event type. Name that file types_by_latitude_by_month.csv.
2. Count the number of event types (snow, rain, etc.) per state per month *for severe events only*. Save these in a CSV file with one column per event type and (50 states x 12 months = 600) rows. Name this file severe_by_state_by_month.csv.

Second transformation: Tables, figures, and listings

These graphs should be generated seriatim by a single driver script that accepts as an argument the directory that holds the analysis files described for each. This driver script should be separate from the one that generates the analysis CSV files described under "data analysis" above, for the purposes of the proof-of-concept.

Please use consistent coloring by event type across graphs, e.g. snow=white, cold=blue, fog=gray, storm=red, rain=green, hail=yellow, other=black

1. Based on types_by_latitude_by_month.csv: generate a table that lists the latitude deciles on the y axis, the months on the x axis, and

the most common type of event (snow, rain, cold, etc.) in the values, along with the count. Color-code each block by event type. Save as most_common_events_by_latitude_by_month.jpg (or whatever graphical format you prefer.)

2. Based on severe_by_state_my_month.csv: generate a stacked histogram that shows the number of severe events by state, for the ten states with the most severe events. Within each bar, show the total number of each type of event, color-coded. Save as top_ten_severe_event_states.jpg (or whatever graphical format you prefer).
3. Based on severe_by_state_by_month.csv: for each state, generate a single graph file that shows a stacked histogram by month that shows the number of severe events. Within each bar, show the total number of each type of event, color-coded. Save as severe_events_by_month_XX.jpg (or whatever graphical format you prefer), where XX = state abbreviation, e.g. (PA, OK, CO, etc.)