US Precipitation Winter 2013/2014

Project Members: Esther Weeks, Jie Zhang, Shinichiro Wachi

Background and Motivation

Winter 2013/2014 has been a winter of extremes. Some areas of the United States have been very wet and snowy, with the most precipitation states have seen in years. Other states, however, have been on the other end of the spectrum, seeing some of the lowest precipitation rates on record for the same period. For example, New Mexico had the driest January recorded with only 0.03 inches of precipitation for the entire month.

Winter precipitation sets the stage for spring and summer conditions including spring flooding, summer drought levels, and even wildfire danger levels; and is therefore important and interesting to a wide range of people. Temperature trends are also of interest, and can reflect climatological changes, as well as affecting plant growth cycles and insect populations for the coming seasons. Furthermore, a series of polar vortices have created the coldest winter in recent memory for people in the mid-west and east coast.

Creating a visualization of this data would provide a quantitative overview of which areas of the country saw large amounts of precipitation and which did not as well as temperature trends for the winter across the country.

Inspiration for this project, was found in the January overview on the NOAA Site which points out some of the extremes.

Project Objectives

Our goal is to provide a way of viewing the precipitation along with snow depth, and temperature from this past winter (2013-2014) in areas of the United States. By providing many ways of viewing the data, we would like to enable people see trends, discover extremes and identify relationships between the weather factors.

Data

The National Oceanic & Atmospheric Administration (NOAA), collects a variety of data from numerous recording stations across the United States. Each station records factors such as precipitation, minimum and

maximum temperature, snowfall, snow depth, and the amount of water in snow. All stations report information daily. Many stations have data records going back into the 1970's. For this proposal, the precipitation, snowfall, snow depth, and temperature (min and max) data will be needed. If we wish to compare precipitation levels between areas that receive snow and those that do not we will also need the measurement of the water content of the snow.

Additional data that would be very interesting to use in the visualization would be the 30-year "normal" measurements or averages for these areas.

Data Sources

NOAA provides the data for this visualization through the following methods:

- NOAA Site, Webservices This option should be able to provide us JSON data from all recording stations. We should be able to filter the data by time, and retrieve either daily or monthly summaries of data.
- NOAA Site, FTP for Data This server provides FTP download of complete data sets. The HCN folder is data from all stations in the US, with all years on record.
- NOAA Site, Text/ASCII for Data This option contains Quality Controlled Local Climatological Data (QCLCD) from about 1,600 US locations. It is downloadable as either ASCII or plain text. As the data has already been vetted by the NOAA, it is easier to work with, requires less sorting and filtering than the FTP data, and is more accurate. Files are available for each month, from around 1996 to the present. Each month's file has the data for all reporting stations. There is an additional file that lists the stations and their information, including city, state, latitude and longitude. Files are easily saved as CSV files in Excel, or similar formats. The sample datasets were obtained using this method.

Additional Data Possibilities

Ideally, having the "normal" or average values for the areas would add a very interesting component to our visualization.

NOAA publishes 30 year average data that we could potentially use for our visualization. These are available here. (Under the chart there is a link to download via FTP). NOAA refers to these as "normals", and they include all of the variables we are already looking at in our data. The difficulty in using the "normals" data is that it is downloaded separately from our primary data and so we would need to match the data we already

have with this new data. Initial exploration into this correlation showed that the station identification number on the normals data sometimes has the WBAN ID(the ID used to identify the stations in our primary data) as its last digits, but either this is not always the case or some of our stations are not on this normals list. If we decide to use normals data, we will have to resolve this conflict or find an alternative. There is the possibility that if a reporting station is listed from the same city as one we are already using, we could use the normals for this, but we need to consider if this is accurate enough for our purpose. Our current proposal is focusing on the data by state, so this might be accurate enough. We would also need to consider how to get this data efficiently. Normals data seems to be downloadable with a separate text file for each station, which could be very time consuming to deal with.

Data Processing

Using JSON data would allow us to have the most current data available up to the current day, however, it seems NOAA is having issues providing this information at the present, so it may not be the most accurate or reliable method. FTP download requires more processing of the data, as it has more information there than we need. The Quality Controlled Local Climatological Data (QCLCD) is easy to download and save as a CSV file. It should also be more accurate and requires less processing as it has more of the data we already need. If we decide to display the data for precipitation, snowfall, and temps by state, we will need to process the totals and averages for each state, as there are multiple reporting stations per state. Assuming we use the last method of data collection, we can anticipate the following:

Pros:

- Lots of data sets, so we should be able to get exactly the data we want.
- The data will be fairly current and widely collected and therefore, relatable to many people.
- There are multiple formats in which the data could be presented.

Cons:

- Large datasets, may take time to sort and filter.
- Need to be specific on data choices; Do we sort by state? If so, there are many stations for each state, do we use all of these?
- Need to be careful to match units of measurements when comparing data. For example, we cannot directly compare inches of precipitation to inches of snow, but could compare inches of precipitation to inches of water in melted snow. (Data is available on chart).

Visualization Proposal

Our goal is to create a visualization that has an overview of the data, as well as the ability to filter down the data to specifics, show trends, explore connections and make comparisons.

Given the data there is an obvious potential to display as a map layout or overlay, but could also be parsed into other (sorted) graphs/charts. A combination of both an overview of the data, as well as additional graphs/charts would be most effective in fulfilling our aims, and allow the viewer good exploration options.

The overview of the data could be represented either by an actual map, of the United States, or as a word tree, where the size of the name of the state and color would depend upon the data, and state names could be placed in their approximate geographical position. The overview map could be filter by temperature, precipitation, or snow, representing one variable at a time. This could also have a slider below filtering by time, that would update the overview to show the data from that time period.

Supplementing the overview would be additional comparisons for a more in depth exploration of the data. These would show data for more than one variable at a time. Examples include comparing temperature to snowfall, temperature to precipitation, or snowfall to water content.

An additional idea for a comparison is to display a scatter plot below our overview, on which the data points are placed relative to the data's location, meaning that points to the right of the graph are from the eastern US, while point on the left are from the Western US.

Many of our ideas also include using the normals value. If it is possible to use this, and we have time, this could be an important addition to our visualization. For example, it should be possible to show how various states accumulate precipitation over time and/or compare this to their normal accumulation rates as a map overlay. More specifically, a state might go from light red to dark blue as a snowstorm hits and pushes the precipitation amount over the normal value. Additionally, an average line could be displayed on a scatter plot, thus making it clear which data points are above or below average and by how much.

Must Have Features

- Overview of data and detail comparisons
- Comparisons of multiple variables
- Geographical display of precipitation / snowfall / temperature / humidity by states
- Timelines (interactivity to select the timepoint)

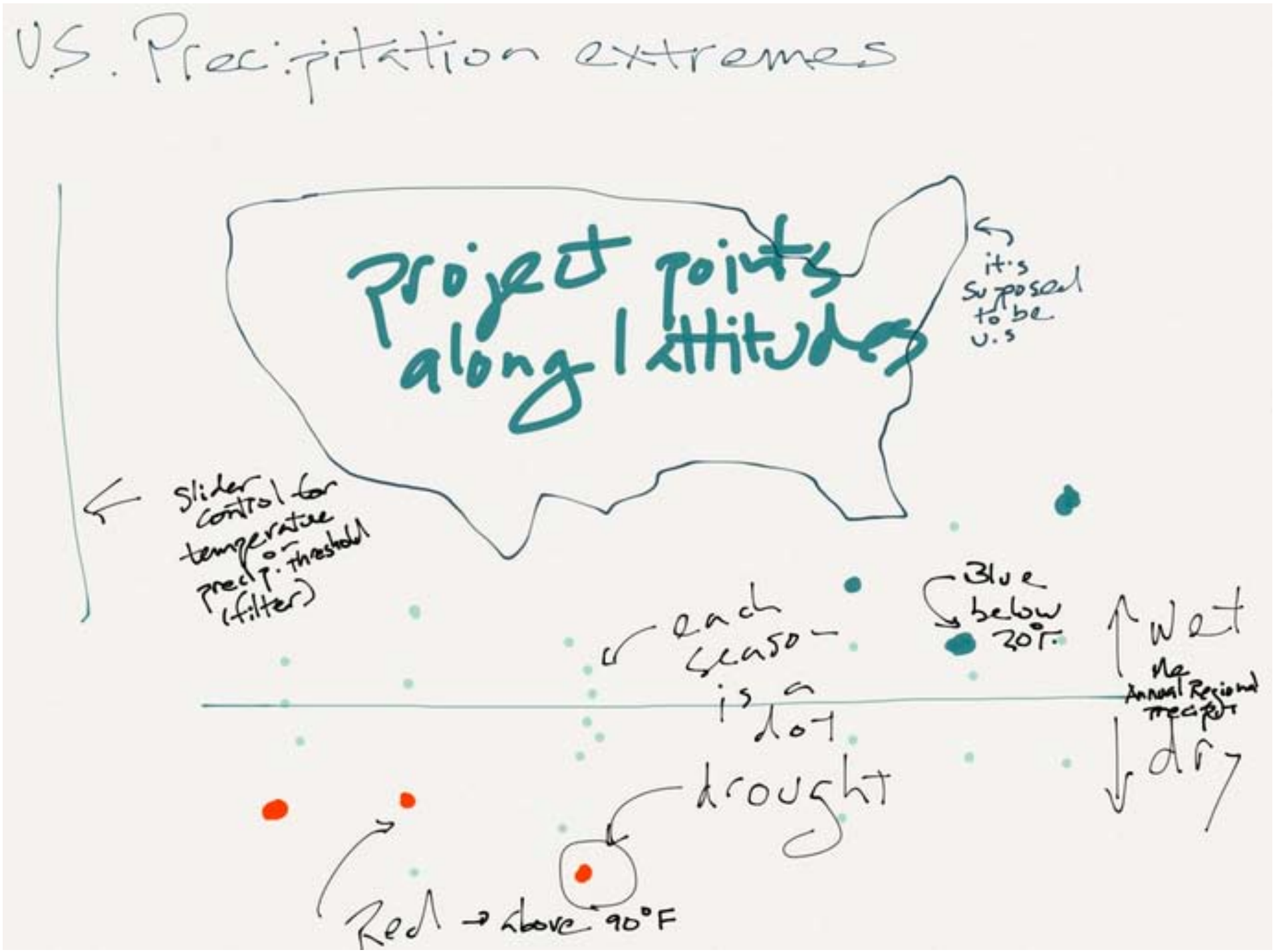
Optional Features

- Normals/Average comparison
- Precise geographical location of the stations (voronoi partitioning of regions?)
- Latitudinal / longitudinal summary plots
- Event maps (mention interesting historical events related to weather extremes)
- Twitter sentiment association with temperature changes (we should be able to collect past tweets based on keywords, such as “hot” “cold” “wet” “dry” from trial service such as Topsy) as indicators for the personal impact of the weather extremes

Project Schedule

Proposed Schedule

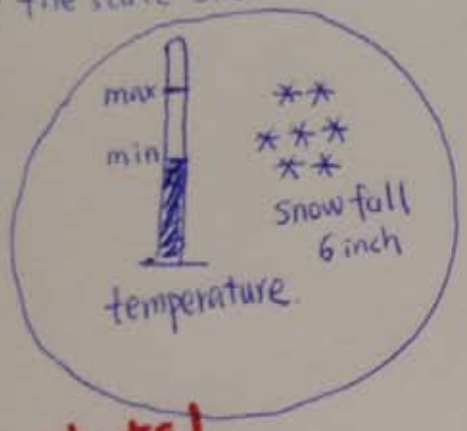
- Setup development environment, website, and process book, etc. (week of March 16?).
- Cleanup and prepare data, look at data with some BI tools like tableau.
- Finish first draft of design and start implementation.
- Implement the design, complete prototype, perform project review with TFs (week of April 14).
- Revise design and implementation based on feedback.
- Create screencast for presentation (week of May 1st).


U.S. Precipitation extremes
Project points along latitudes
it's supposed to be U.S
slider control for temperature or precip. threshold (filter)
each season is a dot
Blue below 20°F
Wet
Dry
drought
Red → above 90°F

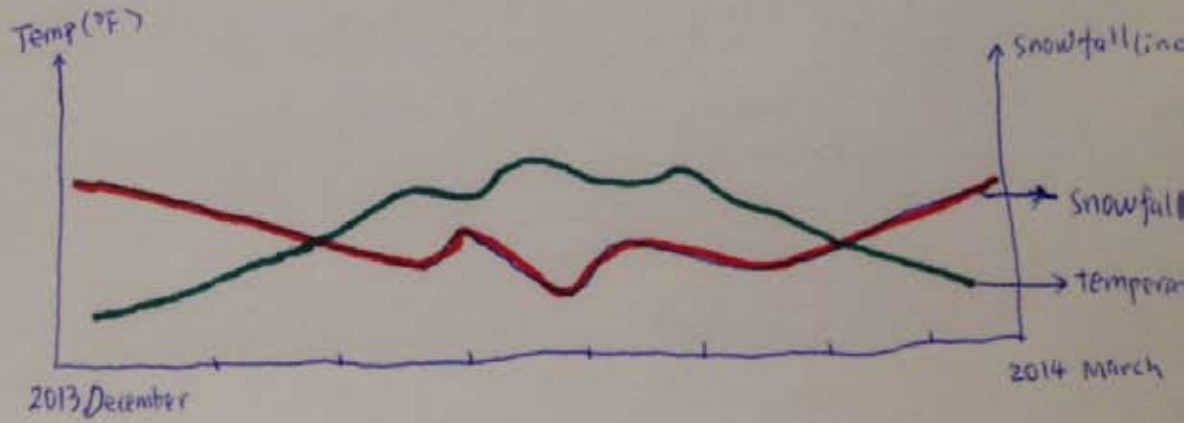
Word cloud formed by statenames. the size and color of each state name is caculated by percipitation dat

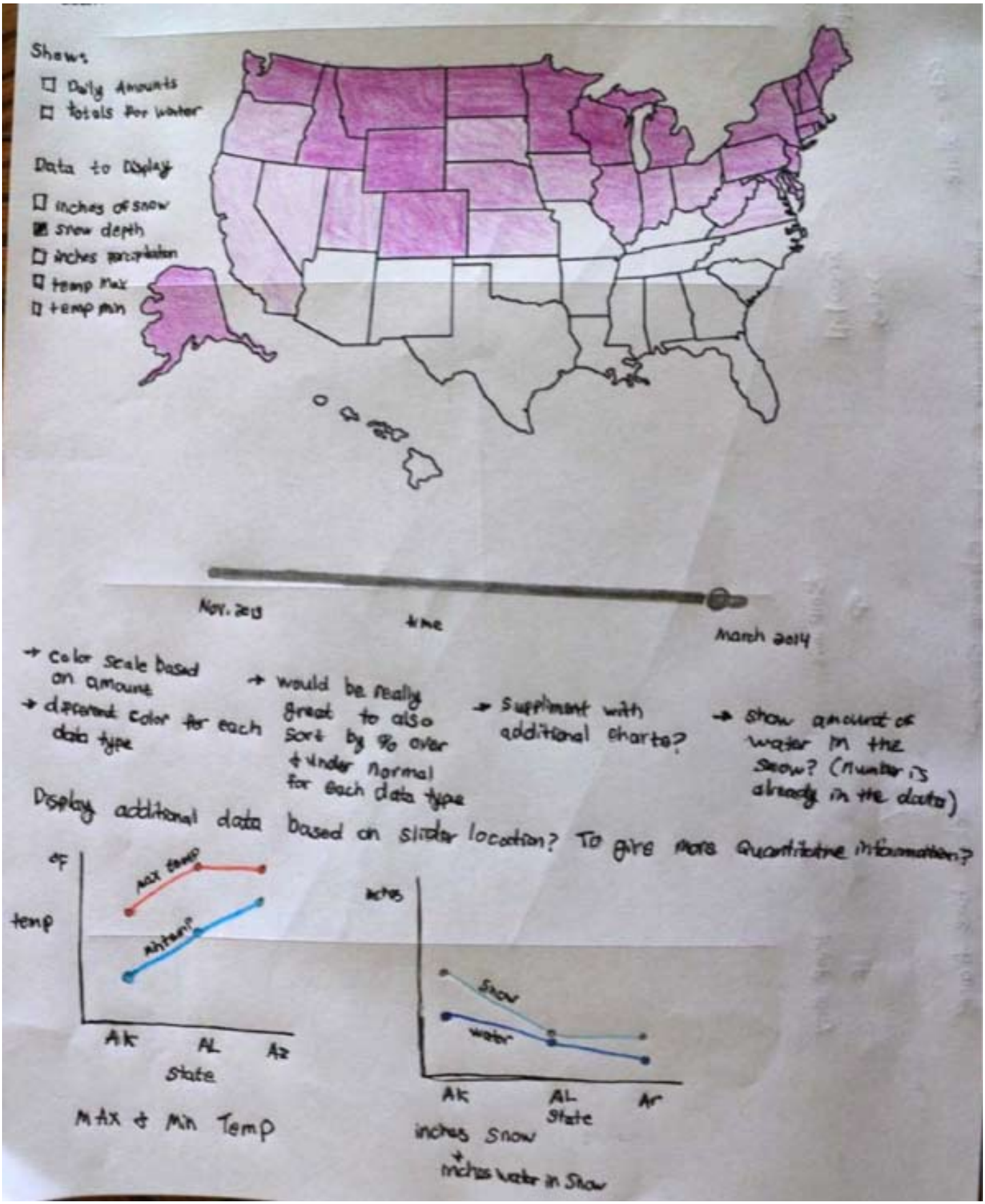
ALASKA Montana NorthDakota Vermont Maine
NewHampchire
Washington Idaho Wyoming South Dakota NewYork Massachusetts
Oregon Utah Colorado Nebraska Lowa Minnesota Illinois ohio Tennesee Virginia
Konsas missouri North Carolina
California Arizona NewMexico Oklahoma Arkansas Mississippi Alabama Georgia SouthCarolina
Texas Louisiana

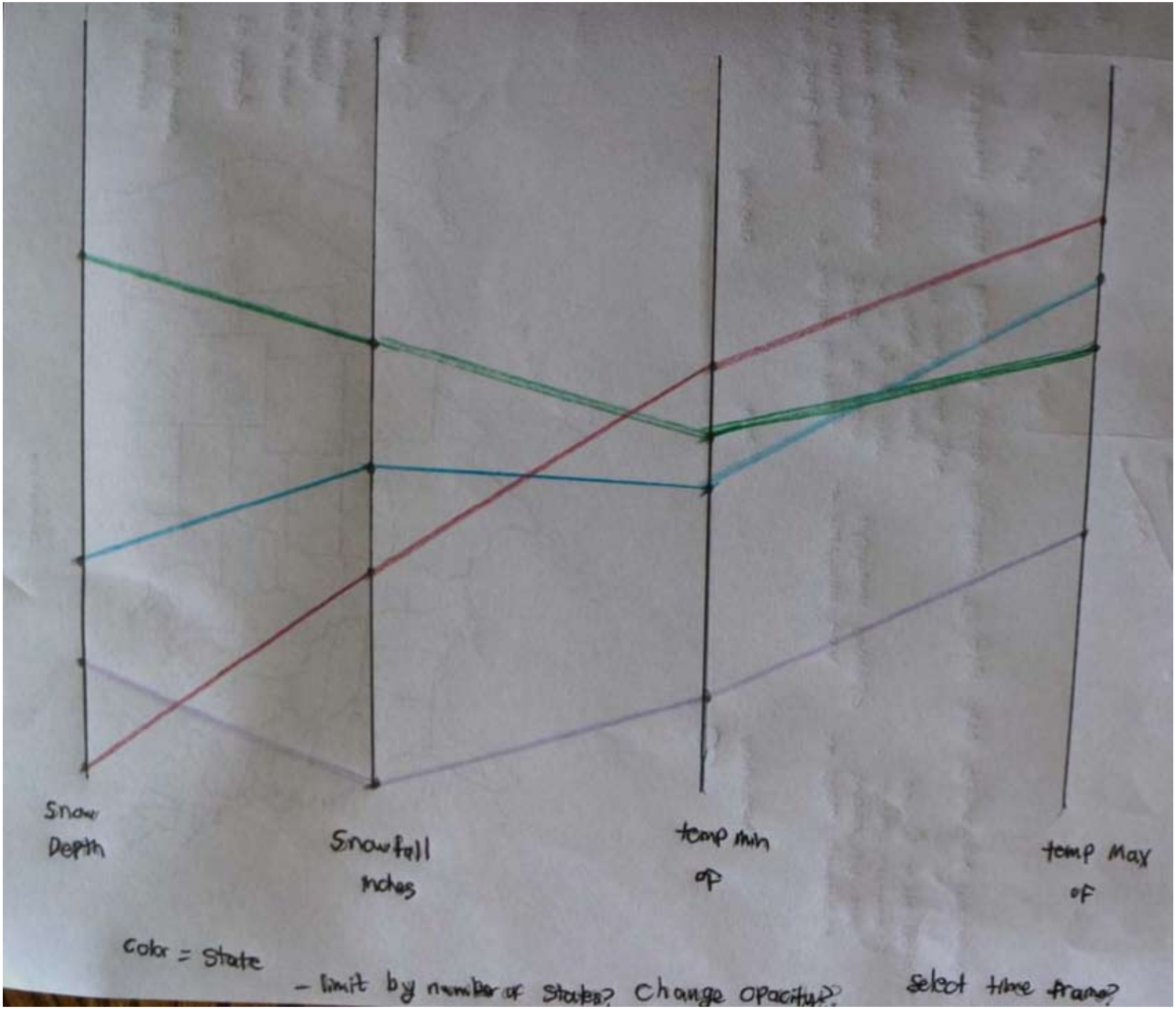
on click of a state, the detail of the state shows below:



total




Shows
Daily Amounts
Totals Per Water
Data to Display
Inches of Snow
Snow depth
Inches precipitation
Temp max
Temp min
Nov. 2013
time
March 2014
→ Color Scale Based on amount
→ different color for each data type
→ would be really great to also sort by % over & under normal for each data type
→ Supplement with additional charts?
→ Show amount of water in the snow? (number is already in the data)
Display additional data based on slider location? To give more Quantitative information?
°F
temp
max temp
min temp
Ak
AL
Az
State
Max & Min Temp
inches
Snow
water
Ak
AL
Ar
State
inches Snow
&
inches water in Snow


Snow Depth
Snowfall inches
temp min °F
temp Max °F
Color = State
- limit by number of States? Change Opacity?
Select time frame?