Southern New Hampshire University

9-2 Final Project: Colorado Storm Data – A Presentation to the State Government

Adam M. Lang

DAT-530-X5611 Present & Visualize Data 19TW5

Professor: Bindu George, Ph.D.

13 October 2019

**I. Background and Audience Analysis**

**A. Background information and research question**

Colorado is a state known for its environmental beauty and home to the Continental Divide of the Rocky Mountains. With the remarkable environment comes some of the most varied weather in the country. There are no months in the state of Colorado without inclement weather and the risks that come with it (See Appendix A, Fig 1). The state Government therefore is going to undertake a data analytics initiative to examine the varied weather conditions of Colorado and how they affect the population from a public safety perspective, risk assessment, emergency management resource allocation, and emergency funding for adverse events. The data available for review is the 2015-2016 storm event data extracted from the [National Centers for Environmental Information Storms Events Database](https://www.ncdc.noaa.gov/stormevents/). This is an extensive data set that includes every weather related event over the 2015-2016 time frame. It includes exact dates and times of events, the geographic coordinates, crop damage, property damage, injuries and deaths.

Initial view of the dataset shows us that there are 23 weather events that occur annually in the state of Colorado. Avalanches, winter storms, flash floods/flooding, lightning strikes, thunderstorms and tornados are the most common events with Hail storms being THE MOST common of them all (See Appendix A, Fig 2).  Each event causes injuries, deaths, property damage and other related risks. May through July are the most common months affected by weather, followed by the winter months of December through March (See Appendix A, Fig 3). Hail is the most common storm type throughout the spring and summer (See Appendix A, Fig 2.). During the winter months winter weather is the highest concern (See Appendix A, Fig 3). Based on an initial assessment-using Tableau, it appears that the data set is sufficient to address the concerns of the state government regarding inclement weather risks.

However, there are a few concerns that may need to be addressed regarding null values in the data set, mostly in the "begin date" and "end date' column which may or may not effect the final data presentation. We do have sufficient data on months that the events occurred which should be enough to make up for the null date data. There are data entries that are not linked directly to a physical location, which may make it hard to detail which geographic areas of the state are most affected by the weather events. There also appears to be a lot of extra language used in "event details" that may be hard to extract and visualize in the manner that the government is looking for. We can extract and visualize the individual weather events, injuries and related events, and the geographic areas but if there are important details missed in the event detail column that may play a role in the final presentation of the data.

**B. Compare and contrast the needs and sophistication of the target audiences**

**i. Message to each audience**

The state budget office is looking to establish a monetary reserve and improve their allocation of emergency resources for storm events in the state. They will be most interested in knowing how many of each event occur (most common and least common), what time of year they occur and what resources need to be allocated for assistance based on any injuries, deaths or natural disasters that occur. Furthermore, they will want to know what geographic areas of the state need the most resources so it will be very important to identify this with the geographic information that we have including physical longitude/latitude coordinates and addresses. A PowerPoint presentation has been requested and will need to focus on presenting this information likely in bar charts of the different storm types and months of the year affected, as well as maps of specific locations effected.

The state's emergency management agency is more concerned with risk assessment, emergency management resources and responding to media inquires regarding relaying information about storm damage to the public. The message will be similar but they are requesting an interactive dashboard, which will require a live up to date information pipeline. The presentation of the dashboard again may be best with graphs, charts, and maps. I think the focus will need to be on what risks are most apparent at what months during the year for the most common and even least common weather events.

Lastly, the state is requesting public safety information in a pamphlet/web page form. This audience will need information presented in a different manner than the state budget office and the emergency management agency. The public needs to know what storm damage can be expected from past data that we have and how to prepare. They will need to know what areas are most affected and what months to expect certain types of high-risk storm weather. Pictures, diagrams and maps may be the most effective way to communicate public safety concerns for storms more so than graphical visualizations.

**ii. How familiar with message at hand will each audience be?**

The state budget office may already be aware of their monetary reserve for emergency resources for inclement weather. What they may not be aware of is the full schema of data behind this monetary and resources reserve. It is important that they understand the level of storm threats, destruction (deaths, injuries, property damage) that is more than likely expected based on this data. It is particularly useful to emphasize the months that are at highest risk for each type of most common and least common disaster. They will need to set aside money for the most common weather events and even the least common that still occur. They may need to ask for more money depending upon our data analysis.

The state's emergency management agency again may be familiar with certain risk assessments, emergency management resources and media inquiries during storms. What they are probably used to are the constant queries from the public and the media regarding management of storm damage and pending inclement weather hazards. If we can provide them with strong data analysis in the form of a real-time working user-friendly dashboard, this may help everyone communicate information at a moments notice in a more efficient manner. This may even prevent them from having to have so many frequent press conferences during storm damage if we can have this valuable data "what to expect" already within reach.

Colorado residents may not be too familiar with public safety information for each type of weather hazard or storm damage. They may only be aware of what they hear on the media coverage and news stations at the time. Giving them access to information that is based on this data set will only increase their awareness, safety and avoid disasters with pending inclement weather during all seasons. I think it will be important to let them know what specific weather alerts to be cautious of, how to prepare for potential storm damage, and what to do if hazardous weather does interfere with their travel, home, business or seasonal activities.

**iii. Educational level, age levels, primary language, cultural diversity, vary among target audiences?**

Considering both the state budget office and emergency management agency I would be more concerned with their level of education and understanding with weather information. There may also be some discrepancies in geographic data that they may not be up to date on. Being that state offices are located in the capital city of Denver, they may not be too familiar with other major and minor geographic centers in the state that are affected by different types of weather events. It will be important to put information in the right context, and make sure that there is no deception in the visual presentation of the data to each audience (Pandey, 2015). A PowerPoint for the state budget office and an interactive dashboard for the emergency management agency are both powerful delivery tools but only if the right visualizations are depicted. The wrong visualization with either too much or too little information can skew the viewpoint of these public officials if they only have a baseline understanding of the information. It will be important to drive home the important points in regards to where resources, time and money should be allocated the most with room to spare if there are weather pattern changes.

The public will need a general education on public safety recommendations for storm damage and resource planning. I do not expect the public to know what we are presenting to them so it will be important that the pamphlet/webpage is a basic run through of important points and visually draws the public in without causing any misleading or deceptive ideas about the information we are trying to convey (Pandey, 2015). The public may include people that are traveling to Colorado on business or vacation and they may not be used to some of the adverse weather events that can and do occur so we need to think about this as a general overview.

**C. Explanation of why each presentation method is appropriate for each audience. What are common objectives and considerations of the target audience? Include communication objectives and considerations of the target audience.**

Each visualization medium is appropriate for the group requested as long as certain variables are considered in the delivery process. The general message to each group needs to be that: Hail is the most common weather event in spring/summer; winter weather is most common in winter; the most storm damage is caused by flooding, hail and tornados; the most deaths and injuries are caused by avalanches and lightning (See Appendix B). Each individual group will need a different message delivered regarding this information. We will also need to educate each group on the least common but dangerous weather events.

The state budget office is going to receive this information in a PowerPoint, which will be a general overview of the above information presented in graphs, charts, and maps. We need to emphasize the most common weather, the types of storm damage, the injuries/deaths that often occur, and the geographic areas of concern. The PowerPoint will be able to deliver the information as a “story” and describe the variability in weather events, and that the most common weather events are not always responsible for the most damage and injuries/deaths (Yau, 2013). They will most importantly need to understand the financial demand for these resources and events and the difference between “what is” and “what could be”. This presentation style will be an excellent way to summarize a call to action of resources and finances for the good of the people (Duarte, 2012).

The state’s emergency management agency (EMA) has asked for an interactive dashboard. An effective dashboard stimulates the audience with: awareness, reflection, sensemaking and impact (Verbert et al., 2013). The main purpose of a dashboard is to bring a sense of reflection and stimulate new ideas with the changing conditions the dashboard presents. I believe we can deliver the proper risk assessments, resource allocation, and media inquiries the EMA is looking for with this medium. Knowing that avalanches and lightning are the leading cause of death; and lightning, thunderstorms and tornados are the leading causes of injury they can communicate to the respective state emergency medical groups how to improve their resources and management. Knowing that damaged property and crops are caused mostly by: flooding, flash flooding, hail and tornados, is vital to allocating storm recovery teams. They will also need to know how to distribute their resources evenly throughout the year since Hail is most responsible for storms during the spring/summer and winter weather is more common in the fall/winter. They also need to be prepared for the least common weather events that still cause damage and require significant resources. Media will be able to easily interact with the dashboard and answer a number of their weather concern questions, and the dashboard will stimulate new thoughts and ideas for the EMA and everyone who uses the dashboard to improve safety during storms.

Lastly, the general public should do well with the pamphlet/webpage. Everything we communicate to the state budget office and EMA in terms of types of seasonal weather, risks and damages that will occur should be displayed but in a broader presentation. General information about public safety using diagrams, pictures, maps and charts will save lives and prevent un-necessary action by first responders and the EMA.

**II. Data Visualization Strategy**

**A. Discuss recommended platform/medium for each audience and how it is going to be used in implementation?**

The three audiences that will be presented with data analytics and visualizations we have to remember are non-scientific as well as decision and policy-makers for the government. Therefore they are presumed to be a highly diverse audience and not always scientific or statistical experts (McInerny, 2014). Government policy audiences need to be able to re-communicate information to secondary audiences (e.g., other policy audiences, companies, public, and the media) and this re-communication of the information delivered to them could be included in the design of visualizations to minimize the biases arising through the chain of communication. The visual interfaces that will be used for these presentations are 2-D and may present challenges for displaying, communicating and understanding multi-dimensional information (McInerny, 2014). The process of visual encoding (understanding the information presented) may only occur by the audience interacting with the data first hand. The problem with 2-D presentation of scientific information is that they convey 'explanatory figures' rather than 'exploratory interfaces' where the audience can learn by doing. The full message may not be fully understood or "decoded" as a result. A way to combat this problem is to create a narrative that focuses on scenarios and data content with a storyline that guides the audience to the important points of understanding.

The goal of these visualizations will be to not only convey a narrative to the audience but to draw the audience attention to important information, to help them organize the information quickly and give the audience a clear structure to follow (Grainger et al. 2016). There are two principles of visualization that need to be followed when using these visual mediums. The first is the "Gestalt [psychological theory](https://www.sciencedirect.com/topics/computer-science/psychological-theory) and associated grouping principles" (i.e. [proximity](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/proximity), similarity, enclosure, continuity and connection) determine some of the ways in which the human brain aggregates objects together (Williams, 2008). The second is the "C.R.A.P. Principles” which involves Contrast, Repetition, Alignment and Proximity (Williams, 2008). The overall goal with these presentations should be to display contrasting quality through graphical features (such as Color/Hue, Size, Orientation, Shape), which will make certain aspects of the visual displays appear to ‘pop out’ from the screen or page ([Wong, 2010](https://www.sciencedirect.com/science/article/pii/S1364815216305990#bib138)). This will lower cognitive load, enhance [pattern detection](https://www.sciencedirect.com/topics/computer-science/pattern-detection) and perceptual inference, and ultimately help the audience interpret and understand the problems in a more logical manner.

The PowerPoint presentation for the state budget office will need to convey salient points that lead them to the direct reasoning behind their investment in a monetary storm reserve and emergency resources. PowerPoint can be an effective medium because the decision-making benefit of PowerPoint results from "both deliberative and heuristic processing (“slow” and “fast” thinking, respectively" (Kahneman, 2011).  There is some fault with PowerPoint though.  Tufte found that PowerPoint’s default settings lead presenters "to create bulleted lists and vacuous graphs that abbreviate arguments and fragmented thoughts" (Tufte, 2013 and Moulton, 2017). Our goal with PowerPoint will be to engage the audience in the level of storm threats, destruction (deaths, injuries, property damage) that is more than likely expected based on this data. It is particularly useful to emphasize the months that are at highest risk for each type of disaster as they may be concerned about the fiscal year of their budget allocation from the state government. They may need to ask for more money depending upon our data analysis. Based on what we know about our audience being non-scientific and the need for them to communicate this information to others in the government, the information should persuade them without heuristic biases but also generate a level of self-understanding. Bullet points with different typography can be an effective way to do this as well as forming an effective narrative. There needs to be an appropriate contrast between graphs and maps that stand alone and offer contrast, repetition, alignment and proximity. The answers need to "pop out" to the audience at a baseline level of understanding (Grainger et al. 2016) I see the implementation as being no more than 10-15 slides: the months with the most weather events, the storm events that cause damage to crops and property, storm events that cause injuries and deaths, the geographic areas of concern, and a summary slide tying it all together with the need for monetary reserve.

The interactive dashboard for the states emergency management agency (EMA) will be easier to convey the information. For implementation we will follow similar guidelines as for the PowerPoint presentation. We will have 5 essential ideas that will be visualized: the months with the most weather events, the storm events that cause damage to crops and property, storm events that cause injuries and deaths, the geographic areas of concern and filters to show the correlation of all graphs/plots/maps together based on months and storm event type. The interactive dashboard will provide real-time data analysis and information presentation. It will also allow decision-making, awareness, motivation and learning (Sarikaya, 2019). The focus for implementation will involve 3 levels of graph literacy: finding and extracting data (Level 1); integrating data and finding relations (Level 2); and analyzing implicit relationships, generating, and predicting (Level 3) (Galesic et al., 2011). There may be a simple learning curve with using the filtering aspect of each dashboard (as there is with most dashboards) but we will provide an introduction so that the users will use the process of self-discovery to form their own narratives about the data.

This is the difference with a dashboard in that it allows “self-driven learning and analytics”, and the data we present must provide that for the users (the EMA). We will also want each interaction between the data filters to have similar to the PowerPoint, a "pop-out" effect where the magnitude of each graph or map that is presented displays the most important relationships (and least important) necessary for them to draw conclusions. For the EMA we will want to focus on geo-visualization using a map to show where the weather events are taking place so they know where to distribute their resources. We also want to focus on the storm damage and injuries/deaths that are commonly seen as they will need to be able to plan to alert first responders and the media. The plan will be for each month where there is weather to drive the dashboard and when you filter/change the month and weather event you can see what are the most common/least common events for damage, injuries/deaths and their locations.

Lastly, the pamphlet or web page will be similar in implementation to the PowerPoint. The audience is different in that it is the general public. Their level of understanding we have to assume is more basic than the previous two audience groups. The problem with the pamphlet or web page is that much like the PowerPoint, it is a 2-D medium and multi-dimensional data may not come across fluently. The design space also may be limited like a PowerPoint slide and this may push us to use bullet points or stand alone visuals/maps/charts which may produce "broken thoughts or ideas" and could leave the audience with confusion if not clearly communicated (Tufte, 2013). We will need to provide a form of a narrative for the audience to have the context of the public safety information that we are presenting (McInerney, 2014). We can visually show them simple graphical and map features (such as Color/Hue, Size, Orientation, Shape), which will make certain aspects of the visual displays appear to ‘pop out’ from the screen or page (Wong, 2010). This will lower cognitive load, enhance pattern detection and perceptual inference, and help the general public interpret and understand the public safety information in a logical manner (Grainger et al. 2016). We also want to focus on simple typography that will deliver the data type intended (literal, categorical, quantitative). Text is essentially used as a qualitative or quantitative data variable and can be interpreted as either based on its use (Brath et al. 2016). Text also is classified by color, orientation, size, shape, position, intensity and containment usage (Brath et al. 2016). If we can use the text to deliver the message with greater size, color and intensity to explain the map or graph presented this will make the "visual decoding" process easy for the audience and they will leave the image with the intended message engrained in their mind.

I see the implementation as being again, no more than 5 main topics, the same as for the previous 2 audiences: the months with the most weather events, the storm events that cause damage to crops and property, storm events that cause injuries and deaths, the geographic areas of concern and associated public safety concerns for each event, storm damage and geographic area. The delivery may be similar to the interactive dashboard but without the interactive component as it will be a 2-D presentation. The interactive component will the be stand alone graphs, charts, maps and the text that "pops-out" and draws the audience to the important information.

**B. For each audience evaluate the types of visual displays (plots) to be used, and if applicable their order (i.e. for PowerPoint presentation) and/or arrangement (i.e. dashboard display)**

When considering what plots to use for each audience it is important to consider what purpose (message) is going to be delivered and the data (variables) to be displayed. I will consider the classifications presented by Vandemeulebroecke et al. 2019 and Reproduced from Margolskee et al. 2018. (Fig. 1 and 2. Below)

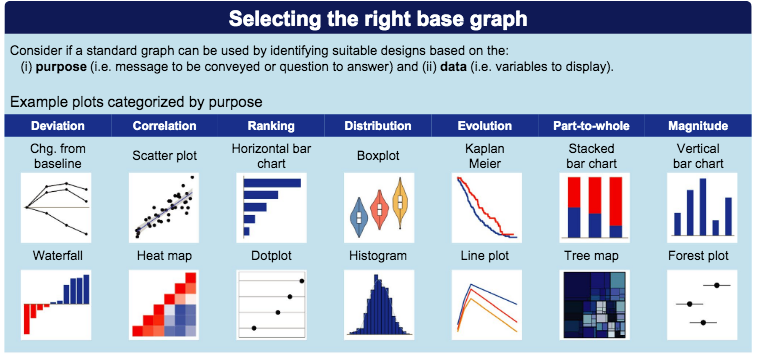


Fig. 1. Classification for selecting a graph for delivery of data.

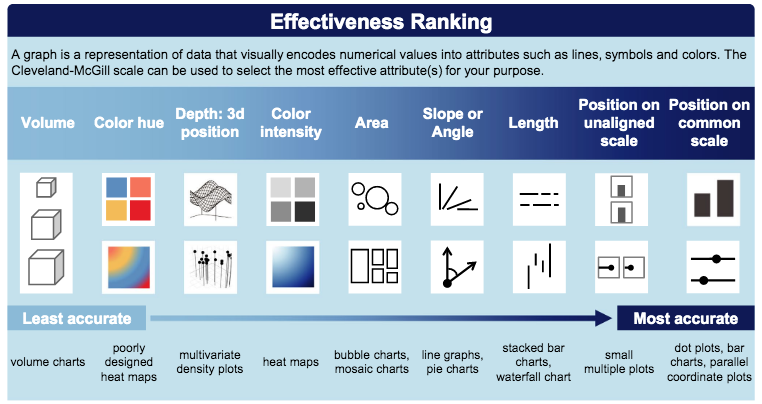


Fig. 2. Classification for selecting the type of graph/plot based on effectiveness.

For the first audience, the state budget office I will use the following outline and plots:

* Horizontal bar graph to show which months produce the most inclement weather in Colorado. This will have a ranking effect and show position on a common scale as seen above. The purpose of the slide is to emphasize which months are of most concern for their budget(s).
* The next slide will present a stacked bar graph to show “parts of a whole” how many months each weather event occurs will be the stacked bars, the x-axis will be the weather events and the y-axis will be the count of the weather events. This will give the affect for example how many months out of the year hail is present compared to winter weather.
* The next few slides will show the weather events themselves again in horizontal bar graphs for a ranking effect, that show which events are most common and least common in each season of the year (i.e. hail is most common in summer and spring, winter weather is most common in fall and winter).
* For the next group of slides I will be showing the damaged crops/damaged property and horizontal bar graphs again may be appropriate to show a ranking effect of which events cause the most damage. Flooding is the main event that will be depicted here.
* For injuries I will show a vertical bar graph with all the weather events on the x-axis and the count on the y-axis. This will show the magnitude of the injuries. Lightning is the top cause of injuries and this will be emphasized.
* For fatalities I am going to take a different approach by using a tree-map, which will show parts of a whole. There were not a lot of deaths in total, only occurring from avalanches, lightning and thunderstorm wind. Therefore we want to show parts of the whole and avalanches and lightning are the two main causes that will be depicted.
* For the final slides I want to show geographic data of location and events. A map of Colorado with bubbles would be appropriate to show area of events. I may also consider using a heat map to show correlation of events with each geographic area that produce the most inclement weather events. The heat map would show the proportional area/size of the inclement weather events to each geographic area whereas the geographic map would not show proportion (Grainger et al. 2016). This is something to consider – and perhaps a heat map would help encode the significant proportion of monetary reserve that is most important for the locations with more events.
* The final slides will have an emphasis on the need for monetary reserve and support for these events. I may consider Tree map or stacked bar graph to show parts of the whole where the most money needs to be considered.

For the second audience the emergency management agency (EMA) I will use the following outline and plots for the interactive dashboard:

* The overall design of the dashboard will be one full screen window with the plots/graphs/maps displayed in 2 rows. The top right of the window will have interactive filters: one filter for weather events and another filter for months of the year.
* The top left corner will be an interactive map where you can click on each weather event and the events will pop up in bubbles on the map thus showing correlation. You will also be able to scan over each bubble on the map and it will show the longitude/latitude, event count, and the weather event associated with the location. The map will have color intensity similar to a heat map with darker colors showing more events.
* Similar to the PowerPoint, to the right of the map will be a stacked bar graph with all storm events. The stacked bars will be months of the year each event occurs, and the x-axis will be the weather events and the count of the events will be the y-axis. The filters will allow this map to show parts of the whole by filtering for month you can look at just the weather events for that month, by filtering for the weather event you will see how many bars there are for that event, for example hail may have 6 total bars showing it occurs 6 months out of the year. Again the stacked bar graph will have a “parts of a whole” effect as depicted in the chart above.
* The bottom row of the dashboard will have a horizontal bar graph in the left hand corner for damaged crops/property to show ranking effect. The center of the bottom row will be the tree map discussed above showing the deaths that occurred from the three main weather events: avalanche, lightning and thunderstorm wind. The last graph on the far right will be the injuries chart that will be a vertical bar graph to show the magnitude of injuries especially from lightning, which causes the most injuries.
* The key with the set up will be the ability to focus/emphasize and de-emphasize each graph by clicking on the filters thus unlocking internal and external visual cues (Kong et al. 2017).

For the final audience the general public I will use the following outline and plots and make a pamphlet:

* The approach to this visualization will utilize text, color and variability to lower cognitive load, enhance pattern detection and perceptual inference, and ultimately help the audience interpret and understand the problems in a more logical manner (Grainger et al. 2016).
* I think the best way to do this would be to focus on the most common weather events (hail in the spring/summer, winter weather in the fall/winter). I will have 3 main pages to begin the pamphlet. The title page will have visuals of the weather events and a basic introduction. The second page will show a map of where all weather events occur and the horizontal bar graph showing that weather occurs in every month throughout the year. The third page behind this will break down the months with the most weather: hail, flooding, tornados, thunderstorms in May-July; and Winter weather December to March. I will use pictures of each type of weather and use the text as the main cue to the information to lower cognitive load.
* The final back page will be broken down into 3 sections. The top section will talk about Flooding and the magnitude of destruction it has for crops and property. I will show this with a vertical bar graph to exemplify the magnitude and show the causes of flooding as part of the graph. I will use typography to explain public safety tips and detail the data.
* The bottom two sections of this will focus on injuries and deaths. I will use a vertical bar graph to show the magnitude of deaths and injuries for the top 2 events in these categories: avalanches (deaths) and lightning (injuries). I will use colored text to highlight the storm data as it relates to these dangerous events and use red text to headline in bold the public safety tips.
* The idea is to reduce the “cognitive load” of the general audience and allow them to focus on the events, the risks, and the safety concerns.

**C. Categorize Visual elements for granularity and sophistication and relate to 3 target audiences**

The level of complexity (count of the number of data dimensions) can influence the visual encoding of the data visualizations (Steele et al. 2011). Usually visualizations with no more than three or four dimensions of data are most commonly used—though visualizations with six, seven, or more dimensions can be found. The more levels of granularity and sophistication of the visualizations needed, the more encoding and visual properties that need to be accounted for. This includes not just the types of information and data used but also the relationships that exist between them. When considering complexity of data levels, it is important to consider data visualization type (exploration vs. explanation) and message type (informative vs. persuasive) (Steele et al. 2011).

Exploratory data visualization is done with a higher level of granularity or detail as usually one is not certain what information the data contains. Explanatory data visualization is the opposite - it usually has a narrative or storyline with low granularity as the messages are more clear and concise. There is also a hybrid level of complexity that encompasses "exploratory explanation" that involves a level of self-exploration of interactive data visualizations such as a dashboard.

Message type, whether informative or persuasive depends on the data level of complexity but also the audience and their interaction. Informative visualizations seek to educate the audience on a subject in a neutral fashion without having to persuade them to make any decisions or analysis of their own. Persuasive data visualization is exactly that: it seeks to change the mind of the audience through its messages (Steele et al. 2011).

For the state budget office they will need a less complex level of granularity in their presentation. I do think they need “exploratory explanation” and “persuasion” in their presentation. The concrete data will be explained in each PowerPoint slide such as that hail is the most common weather in spring/summer and winter weather is most common in fall/winter. They will also need to know outright what events (flooding) cause the most destruction of crops and property, and what events cause the most injuries (lightning) and deaths (avalanches). The reason for “exploratory explanation” is that we need to give them the weather events/months that cause the most problems – that will be the explanation and we need to persuade them to direct their state funding towards those events. There will be a level of exploration for this audience though. We will also have to show the weather events/months with the least problems and this may require exploration on our presentation part and the audience understanding of where to direct the least funding but still have reserve funding available – what events can they correlate with the events causing the most problems? There needs to be an finite contrast between graphs and maps that stand alone and offer contrast, repetition, alignment and proximity. The answers need to "pop out" to the audience at a baseline level of understanding so the audience can communicate the complexity of the message to persuade their government superiors where to direct funding (Grainger et al. 2016).

The interactive dashboard for the Emergency Management Agency (EMA) will by itself provide “exploratory explanation” as the users interact with the various visual encoding found within the dashboard. They will be able to formulate their own level of analytical understanding of the data. The data will be provided in real-time, which will give them the chance to direct their storm teams and resources in the proper directions.

The pamphlet for the general public needs the least granularity and sophistication of all of the audiences. There will be simple explanation and informative data for them to make proper adjustments for each weather event (Steele et al. 2011). The simplicity of the messages will lower cognitive load, enhance pattern detection and perceptual inference, and help the general public interpret and understand the public safety information in a logical manner without the need for exploratory analysis or persuasion (Grainger et al. 2016).

**D. For each audience describe the formatting guidelines that will be used for effective delivery of the message**

When considering format of the visual design space, it is important the area is used efficiently to convey the intended messages of data (Brath et al. 2016). Visual encoding can be perceived differently based upon the semantics of the font and labeling; color hue, saturation and brightness; and positioning of the visual elements (Brath et al. 2016). Humans automatically process and categorize visual stimuli into regions and properties based on “pre-attentive processing” (Healey et al. 2012). We then often use what is called “ensemble encoding” where with a glimpse of visual data we can extract different information in parallel to one another and come away with multiple messages at once (Healey et al. 2012). The goal of any design should be to direct the reader through proper visual cues to the most important information first. Internal and external cues can be used by simply emphasizing or de-emphasizing font, labels, context, colors and positioning (Kong et al. 2017).

The PowerPoint for the state budget office is a challenge for visual encoding of the intended message of the data. This is because PowerPoint’s default settings lead presenters "to create bulleted lists and vacuous graphs that abbreviate arguments and fragmented thoughts" (Tufte, 2013). Therefore the focus of every slide should be on using typography (text) as a headline that font weight, using size and intensity, can be used for encoding quantitative or ordered data (Brath et al. 2016). We want to immediately direct the thought process of the audience that each slide is going to show the most vs. least weather events for each season, the most destruction, the most injuries/deaths and where the locations are most affected.

Each slide will have a graph or picture on the right hand side and the text encoded on the left hand side. I may have to place certain graphs on their own slide to let the data stand out on it’s own and allow the audience to decode the intended message (Brath et al. 2016). It is known that changing the weight of the text (boldness) as well as using underlining or italics can make the typography stand out to the audience immediately with the intended message; we also read left to right (Brath et al. 2016). So for example on the first slide if we are displaying a stacked bar graph of the most common weather events in spring/summer (i.e. hail) the focus should be on that phrase so they do not have to spend a lot of time visually decoding the graph; they read the text on the left hand part of the slide and their vision is immediately externally cued to the right hand graph, or cued to look for the data in the visualization on the next slide (Kong et al. 2017). The color of the text I may consider a theme of evergreen and brown to give the audience the feeling of being in Colorado with its green trees and brown mountains, and these colors will act as background to the graph and highlighted text delivering the message(s). I would consider using red with underline and boldface to highlight the important text.

The color of the graph bars will also be important. Since we can’t order color hue, but we can order saturation and brightness (Steele et al. 2011), I may consider using a bright color like red to make the greatest weather event in summer (i.e. hail) stand out among the other data points. Red also induces a feeling of anxiety, which would make the audience more persuaded to push more money for the weather event causing the most problems (Bartram et al. 2017). Lastly space is very important in PowerPoint slides as the default settings do not give you many options, often times why presentations are fragmented in visuals and ideas (Moulton et al. 2017). By placing the text on the left and the graph on the right we are using segmented ensemble encoding (Szafir et al. 2016). The user will take in the typography and then parallel and add the thoughts together as they read into the graph on the right hand side. External cues such as contour of the boxes of each side of the slide, as well as labels on the graph will be important (Kong et al. 2017).

One last granular layer of sophistication to consider with the Power Point presentation is the narrative that goes along with it. As a data scientist we usually present the information to the audience. There may be an initial presentation that I give to the state budget office, but it may or may not be recorded, and the bullet points on the slides may not be enough for the state budget office to then communicate the presentation to their superiors in the government. Therefore I would also include notes at the bottom of each slide as a detailed narration to direct each slide if the presentation is given again (McInerney, 2014)

The interactive dashboard presentation for the EMA will take a different approach in use of visual cues. Unlike PowerPoint where the slides will be easily read left to right, the dashboards can be a bit more complex and could confuse the readers as where to look or use the interactive buttons and graphs. For the EMA we will want to focus more on the geo-visualization using a map in the top left corner to show where the weather events are taking place so they know where to distribute their resources. By placing the map in the upper left corner, the reader will be cued by the locations to then look at the other dashboard aspects related to the geographic event. We also want to focus on the storm damage and injuries/deaths that are commonly seen as they will need to be able to plan to alert first responders and the media. The plan will be for each month where there is weather to drive the dashboard and when you filter/change the month you can see what are the most common/least common events for damage, injuries/deaths and their locations.

To format the dashboard, the font will be similar to the PowerPoint slides with font weight, size and intensity directing the messages. The same colors evergreen and brown will be used to give the audience a feeling of being in Colorado (Bartram et al. 2017). I will underline and highlight in red what is most important or severe as red is known to induce anxious feelings (Bartram et al. 2017). The plan will be to use headlines for each dashboard graph or chart. The filters will be on the right hand side and will include months and weather events, then once you filter you will be directed to the graphs that take up the majority of the screen. You will start with the map in the left hand corner and follow it to the storm events, then damage, then deaths, then injuries. By clicking on each weather event you will be able to see a graph load the data that shows for example that avalanches cause the most deaths in winter. For the map part of the dashboard I would really like to use a heat map to show the intensity of where the most weather events occur. In heat maps, the higher the intensity of the phenomena, the higher the intensity of the color and opposing phenomena (negative/positive) have different, two-color tones in a divergent pattern (Netek, 2018). A study on traffic accidents used green for less intense, yellow for medium and red for most intense data (more accidents) to show the contrast in event number and type (Netek, 2018) so I may consider this color schema for the heat maps.

Lastly positioning is again considered important as for the Power Point slides. Since with ensemble encoding we can parallel multiple images as we read left to right and segment the data in categories in our brain, I am going to position the graphs and maps on the left hand side of the dashboard and the filters on the right hand side so the eyes are directed left to right just as if they are reading a book. You will filter on the right, then read the dashboard left to right (Healy et al. 2012).

One last consideration with the dashboard is the lack of a narrative on the screen of the dashboards. The interactive nature of the presentation involves three levels of graph literacy: finding and extracting data (Level 1); integrating data and finding relations (Level 2); and analyzing implicit relationships, generating, and predicting (Level 3) (Galesic et al., 2011). There may be a learning curve with using the filtering aspect of each dashboard, but we will provide an introduction so that the users will use the process of self-discovery to form their own narratives about the data and then teach it to other users in the future.

For the pamphlet the format of data visualization for the general public audience needs to be the least granular. It needs to be a simple explanation of the public safety and awareness to consider with each storm event and possible destruction. Knowing that the average person can’t process more than 5 colors at once, I would keep the same color schema as before using evergreen, brown and red or orange to highlight information (Bartram et al. 2017). The typography for this section needs to be simple, bold-faced without much style or character. The use of underlining and boldness may be necessary to act as internal cues for the reader (Brath et al. 2016 and Kong et al. 2017). I would like to set this format up similar to the PowerPoint presentation with the text on the left hand side of the page, which will direct them to images on the right hand side of the page. Again the text will be bold, underlined or red where a most important data element must be conveyed (Brath et al. 2016). The idea is that they will parallel the encoding of the text to seeing a bar graph or map on the right hand side of the page. I may also consider using a heat map as for the dashboard presentation with the same colors discussed above (green for not so severe, yellow for medium, red for high intensity) so that we can detail the geographic areas for the public of where the most weather events occur. The idea with this presentation is to keep the text and graphics simple. The background will be a light green color keeping with the Colorado theme but also making it so that their internal visual cues are to the information presented not to the art display behind the data (Kong et al. 2017).

**E. Logically assess possible feedback methods for each audience that will uncover possible comprehension or technical problems**

* **How will each audience submit feedback or ask questions about your final solution?**

The state budget office as well as the emergency management agency (EMA) will be able to submit feedback and questions via email, Kanban boards, telephone and inter office mail correspondence. I think it would also be appropriate to set up meeting times with each group so that any type of feedback or questions can be discussed in an open forum. This would be similar to a scrum meeting that a data analyst and their team would have while working on a project with stakeholders (Sliger, 2011). After the meeting there may be new ideas or directions for the project to be taken thus stimulating more data analysis and reporting. This is important to note as not all data analytics projects are set in stone once they are completed. It is important to hear the feedback from the stakeholders and use it for future projects their offices may pursue or we may pursue as data analysts.

The general public can submit feedback via email, phone, web correspondence, or the state government offices. They may also submit questions and feedback via the local media.

* **Assess various methods and discuss any technical or other challenges for each audience**

The state budget office may need to have multiple scrum type of meetings to go over the data presentation. This is because once they present the information to their superiors (i.e. the Governor), the state government (house, senate) may want to meet themselves and more questions may arise regarding why or how they should increase or decrease funding for certain weather related programs.

The EMA may also need to meet multiple times to discuss the dashboard if there are personal analytical findings they come across while using it. As stated there is a learning curve with dashboards, but there may be new inquiries that arise or new requests they may have regarding the data. For example, they may find that they are sending more emergency teams to certain parts of the state more than what the data we are using from 2016 is showing them.

The general public may want to hold a public meeting such as a town hall to go over some of the public safety recommendations. This may even involve some local government officials. A challenge for the general public is getting their message out as to whether the intended recommendations in the pamphlet is actually helping them or not. The local media may again come in to play as they may have questions about how the public is reacting to the recommendations. Again a town hall type of meeting may be warranted and the local media can host this type of event.

Overall a major technical challenge with each group will be the learning curve of the presentation medium (PowerPoint, interactive dashboard, pamphlet). As with any presentation there will be information not included that may come up in discussions or the process of discovery. For example, the state budget office will receive the information regarding the storms that cause the most and least damage/injuries/deaths and their locations. They are supposed to formulate a budget based off this information. I think a major technical challenge will be that the presentation will not have any cues as to how to best estimate their budget. The 2015-2016 weather dataset we are using does not have any monetary storm damage totals just total numbers of damaged property/crops and total numbers of injuries/deaths per storm event. Perhaps there is a way to include this information separately but I do think it would be challenging to do so. This may leave this audience wanting to see an entirely different data set with monetary spending from this time frame of 2015-2016 and we don’t have that information in our reporting.

I think the same type of challenge goes for the EMA dashboard. While we are creating a real-time dashboard in which they can interact and perform exploratory explanatory analytical granulation on the data, again the information doesn't include the number of resources or emergency teams/personnel used during each storm. I would assume that the EMA has a method of estimating from previous years how they delegate their storm teams and first responders and could apply these formulas, but they too may want more information as to see for example how many storm teams they had to deploy for an average tornado or winter storm. The hope is that the information we are supplying them increases the efficiency of their storm operations, but there is no way to predict what information they don’t have until we present them with the data analytics that we do have.

**References**

Bartram et al. (2017) Affective Color in Visualization. CHI ’17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, CO, USA, May 6-11, 2017). Pgs: 1364-1374.

Brath et al. (2016) Using Typography to Expand the Design Space of Data Visualization. She Ji:  The Journal of Design, Economics, and Innovation. Vol. 2 (1):59-87.

Duarte, N. (2012) Structure Your Presentation Like a Story. Retrieved from: <https://hbr.org/2012/10/structure-your-presentation-li>

Galesic et al. (2011) Graph literacy: A cross-cultural comparison. Medical Decision Making. 31(3): 444-457.

Grainger et al. (2016) Environmental data visualization for non-scientific contexts: Literature review and design framework. Environmental Modelling & Software. Vol. 85: 299-318.

Healey et al. (2012) Attention and Visual Memory in Visualization and Computer Graphics. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 18, NO. 7: 1170-1188.

Kahneman , D. (2011) Thinking, Fast and Slow. London, England: Macmillan.

Kong et al. (2017) Internal and External Visual Cue Preferences for Visualizations in Presentations Eurographics Conference on Visualization (EuroVis) 2017 Volume 36, Number 3:

Margolskee et al. (2018) Graphics Principles Cheat Sheet. (Biostatistical Sciences and Pharmacometrics, Novartis Institutes for Biomedical Research, Cambridge, 2018). <https://graph icspr incip les.github.io/>

McInerny et al. (2014) Information visualization for science and policy: engaging users and avoiding bias. Trends in Ecology & Evolution. Vol. 29(3): 148-157.

Moulton et al. (2017) Does a presentation’s medium affect its message? PowerPoint, Prezi, and oral presentations. PLoS ONE 12(7): 1-39.

National Centers for Environmental Information Storms Events Database 2015-2016 data set

Netek et al. (2018) Implementation of Heat Maps in Geographical Information System – Exploratory Study on Traffic Accident Data. Open Geosciences. Vol. 10(1):367-384.

Pandey et al. (2015) How Deceptive are Deceptive Visualizations? An Empirical Analysis of Common Distortion Techniques. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Pgs: 1469-1478.

Sarikaya et al. (2019) What Do We Talk About When We Talk About Dashboards? IEEE Transactions on Visualization and Computer Graphics, 29(1): 682-692.

Sliger, M. (2011). Agile project management with Scrum. Paper presented at PMI® Global Congress 2011—North America, Dallas, TX. Newtown Square, PA: Project Management Institute.

Steele et al. (2011) Designing Data Visualizations. Sebastopol, CA: O’Reilly Media, Inc.

Szafir et al. (2016) Four types of ensemble encoding in data visualizations. Journal of Vision. Vol. 11(5): 1-19.

Tufte, E (2013). The cognitive style of PowerPoint. Cheshire, CT: Graphics Press.

Vandemeulebroecke et al. (2019) Effective Visual Communication for the Quantitative Scientist. CPT Pharmacometrics Syst. Pharmacol. 2019 Jul 22. doi: 10.1002/psp4.12455. [Epub ahead of print]

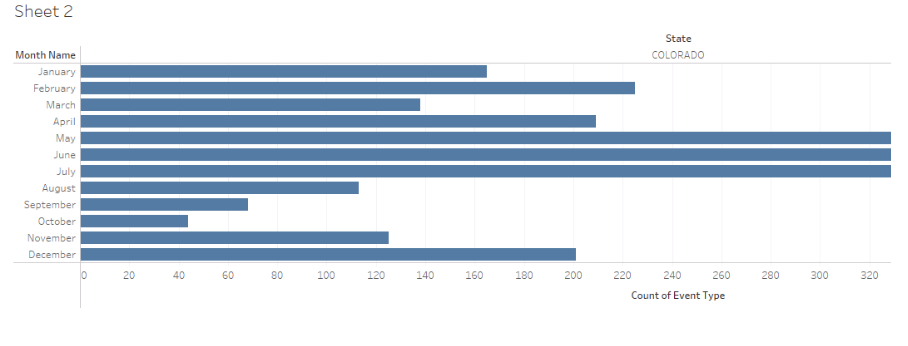
Verbert et al. (2013) Learning Analytics Dashboard Applications. American Behavioral Scientist. XX(X):1-10.

Williams, R. (2008) The Non-designer’s Design Book (third ed). San Francisco, CA:Peach Pit Press.

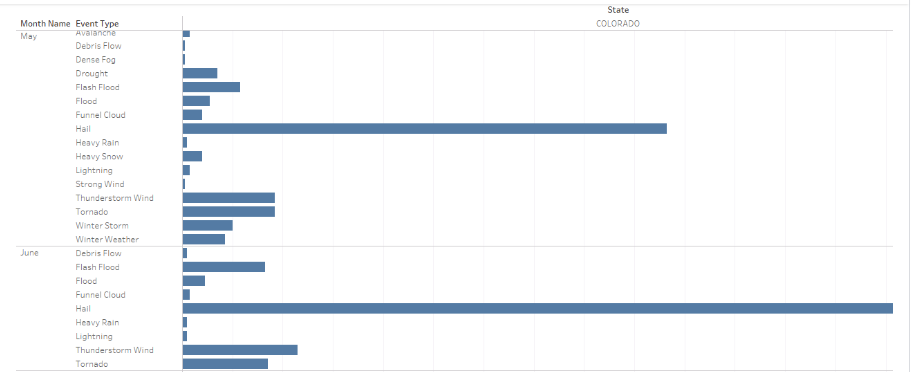
Wong, B. (2010) Points of View: Salience. Nat Methods, Oct, 7(10):773.

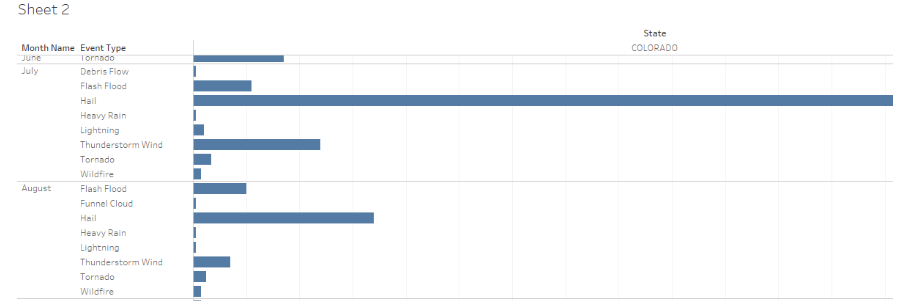
Yau, N. (2013) Data Points: Visualization That Means Something. Hoboken, NJ: John Wiley & Sons.

**Appendix A: Tableau Analysis of 2015-2016 Colorado Storm Data**

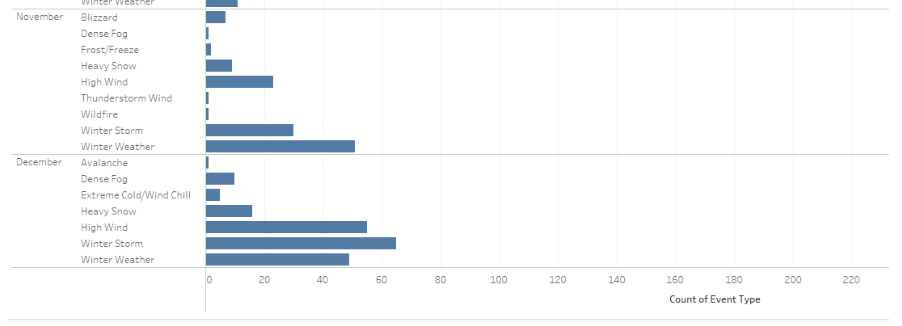
****

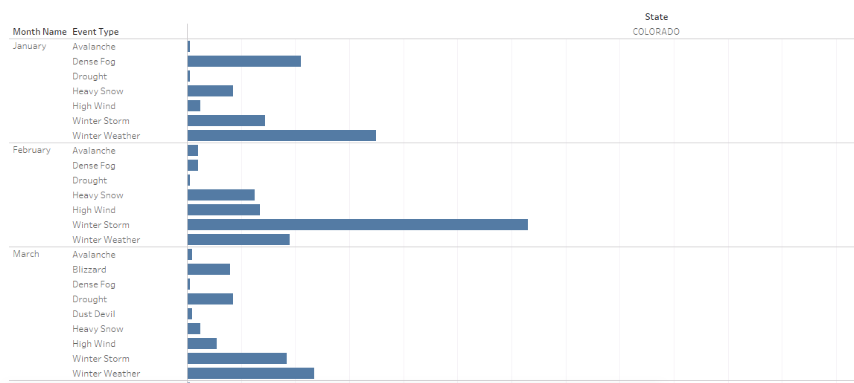
**Fig. 1. Weather events occur throughout the year in Colorado.**

****

****

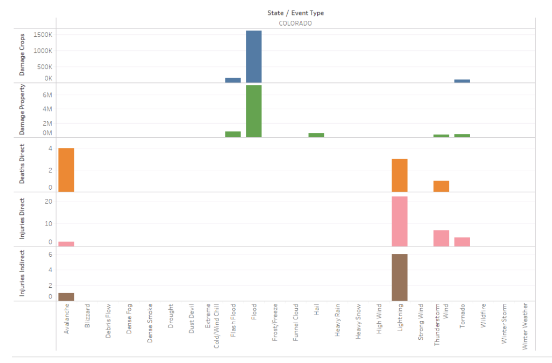
**Fig. 2. Spring/Summer weather for Colorado. Hail is the most common storm type.**

****

****

**Fig. 3. Winter weather is most common in fall/winter/early spring in Colorado.**

**Appendix B: Tableau Analysis of 2015-2016 Colorado Storm Data**

****

**Fig. 1. Visualization part 2 of storm damage, deaths, injuries in Colorado 2015-2016.**