Project Report: Snowplow Data

My capstone data project involves snowplow data provided by Montgomery County Department of Transportation-Highway Services(DOT). Additionally, I will be using datasets provided by the Department of Environmental Protection(DEP) who are responsible for tracking the stream health conditions. DOT provided me with seven different shapefiles that record the snow routes used within the county. These include routes named Silver Spring, Bethesda, Colesville, Gaithersburg East, Gaithersburg West, Poolesville, and Damascus. Moreover, each snow route shapefile contains specific routes used for that depot. Depending on the depot, the records range from 300-2500 with more than 30 attributes.DEP gave me two datasets, one of which contains the actual watershed locations/boundaries, and another which contains information on the health status of streams/tributaries. Please note: I am still missing datasets that track salt/brine usage. However, with the current dataset made available to me, I understand that these datasets come from trustworthy sources such as Highway Services and a biological monitoring group at Montgomery County. During my project, I used two tools: R Programming and QGIS. Although this may be true, I primarily used R for the majority of my project and in conjunction with libraries such as Tidyverse, SF, RGAL, Htmlwidgets, and Leaflet. My goal for this project was to create an interactive web map that my clients could use to identify the distribution/overlay of routes and watersheds. For the purpose of mitigating the impact of salt usage on watersheds and the various ecosystems that rely on clean water.

After ingesting the data into R, there were a few cleaning steps needed in order to perform an accurate analysis of the data. For example, changing the CRS/Projection system to a more readable format to use Leaflet. Meaning, that these datasets came with their own predefined custom CRS which Leaflet had trouble reading. As such, I changed the CRS Coordinate system to 4326. Once this was completed, I needed to rename a few columns to produce a more user-friendly column. I renamed columns “Popup” and “Narrative” into “Info” and “StreamConditions.” Info now contains information on the flow of water in streams and StreamConditions contains information on its health status. The different health statuses include Excellent, Good, Fair, Poor, and No data available. I also assigned each status a color to properly identify these values. Afterward, I had to merge all seven snow route datasets into one major data frame. It is important to mention that this was easily possible given the uniform standard of how all the data was collected and recorded. And so, once the data was merged I needed to create two smaller data frames derived from it which record emergency and non-emergency routes. However, the values contained in this column were “E” and “2”. After a meeting with my mentors, it was discovered that E meant emergency routes and 2 were non-emergency routes. As such, I had to rename all values to their proper meaning for better clarity. Additionally, the column that contained snow names, “Route_name” had underscore values for both Gaithersburg depots. Therefore, I used the gsub function to drop any underscore so that It wouldn't appear in the map. Lastly, I merged the dataset named “streamconditions” with “snow_routes” using the st_intersect function. I did this because I wanted to get a count of each snow depot and how their routes intersected with the tributaries. For example, how many Silver Spring routes intersected with Excellent or poor tributaries? These were the major steps taken to clean and prep my data for a thorough analysis needed for the successful completion of my project.

For my descriptive visualization portion, I created a stacked bar chart. On the x-axis, I had all seven of the snow route depots. Again these were Silver Spring, Bethesda, Colesville, Gaithersburg East, Gaithersburg West, Poolesville, and Damascus. Afterward, I got the count of how many times each of their route segments intersected with tributaries. Additionally, I filled the colors with the statuses of their health. Meaning, that we could see how many routes are distributed into poor or fair areas. As a result, the user would be able to identify which routes are the biggest and smallest in terms of counts and how each route overlays between the tributaries. This is a simple visualization that really gets the point across to my audience and leads the way to my data product.

For my capstone project, I created an interactive web map as my data product. The reason for developing such a product is because knowing the context of where routes are and how they overlay with the tributaries was an essential component of analysis needed for my clients. When I was first working on this project I created a map using tmap. However, this was a static map, meaning the client was unable to get context or make any useful insights. Another reason for creating an interactive web map was to have the ability to toggle on and off between layers. Again, I had seven snow routes that I needed to plot, and having them all plotted at once was really clustered and indigestible for my clients. Whereas, with an interactive web map, the user would be able to switch between routes to better compare and contrast between them. And so, I was exposed to a library in R called Leaflet which gave me the exact tools I needed to mimic the functions and abilities of ArcGIS or other GIS software. I incorporated several key features to enhance its usability and functionality. One of the primary features I implemented was a geolocator tool. This tool allows users to pinpoint their current location on the map accurately. By leveraging geolocation services, users can visualize their position relative to the snow routes and tributaries. This feature enhances user engagement and facilitates personalized analysis based on their geographic context. Another feature I added to this was a marker. Meaning, that users would see their location with a marker and identify the tributaries, their health, the watershed, and the routes, whether it was an emergency route or non-emergency route. I think this is very useful for Montgomery County residents as they could identify which routes pertain to them. To improve user experience and navigation, I included a home button functionality. This button serves as a convenient shortcut for users to reset the map to its default view with a predefined zoom level. Upon clicking the home button, the map resets to its initial view, providing users with a consistent starting point for exploration. This feature streamlines the user interface and ensures ease of use, especially for users unfamiliar with interactive web maps. In my personal case, I live outside the county and when I used the Geolocator feature, I often had to scroll all the way back to find the map. And so, I had the idea to create this shortcut for others who may not be in the county and for ease in resetting the map to its original view. The primary layer is called “Streamconditions” which displays the status of tributaries and streams within the county. Users can click on individual tributaries to access detailed information about their condition, including water quality, stream flow, and name. By visualizing stream conditions on the map, users gain valuable insights into their health. Another essential component of the map is the watershed 12-Dig layer. This is something that my client wanted to include in the map which could help aid efforts in identifying watersheds that are at risk. The map includes overlays of seven snow routes used by the Montgomery County Department of Transportation-Highway Services (DOT). Each route is represented as a distinct layer, allowing users to toggle between them to avoid clutter and focus on specific routes of interest. Users can explore the spatial distribution of snow routes and assess their proximity to residential areas, critical infrastructure, and environmental resources. Additionally, the map features layers distinguishing between emergency and non-emergency snow routes. Users can click on individual routes to view depot names, route identifiers, and other relevant information. This feature helps users identify the routes assigned to emergency response teams and prioritize snow removal efforts during inclement weather events. By integrating these layers into the interactive web map, users can explore a wealth of spatial data related to stream conditions, watershed boundaries, and snow route assignments. The toggle feature enables users to customize their viewing experience, while clickable elements provide access to detailed information, empowering users to make informed decisions. Especially for my clients and Montgomery County residents who wish to stay informed on snow routes and watershed health.

My data story involves the current condition with Montgomery County route data. Prior to this, DOT did not know how routes overlaid with tributaries or watersheds. As such, they were unaware of the potential harmful impacts of salt caused by their routes. Excessive road salt can negatively affect wildlife, streams, rivers, soil, infrastructure, and human health. Especially in urban or highly developed areas where salt begins to combine with other metals and chemicals. Not to mention, that once salt is applied… there is no easy of removing it from the environment. Moreover, the lack of fish and bug life are key indicators of poor stream conditions. This is because sensitive fish cannot thrive in these conditions due to the harsh chemicals. Additionally, tributaries with poor conditions are often the ones with more routes serviced to them. With my data product, the user will be able to identify these areas and begin a course of action to better improve these conditions. For example, limiting salt usage or better management/training. And so, with this new information, we can minimize the amount of salt application that enters streams, rivers, tributaries, and watersheds which can potentially harm the environment. Meaning, that Montgomery County residents will be able to enjoy clean drinking water and wildlife will also have the ability to thrive around the county.

My overall experience working with MC was rather good. I was able to gain exposure to the world of GIS. Meaning, I learned about GIS tools/concepts, day in the life, career paths, and more. I think this was by far one of the biggest pros. Additionally, I was able to increase my knowledge and awareness of the potential impacts of salt on the environment. Knowing that my work has real-world usage and application means something to me. Of course, I'm not solving the issue at hand, but by completing a step in this important process I feel as though I contributed to something meaningful and good. I was also thankful for the help and guidance that I received from my mentos at DOT/MC who helped me throughout this process. They were very patient and kind to me which goes a long way. My team took the time to explain important data concepts and what the data itself reflected. However, there were also some cons throughout the process. Most important, was the lack of knowledge in the GIS space with its tools and concepts. It was a very huge learning curve that took a significant amount of time to learn. Additionally, not being able to use ArcGIS which is the standard tool to use in the GIS space was unfortunate. Especially given that I was very limited in the amount of tools and analyses that I could do. Lastly, another con was the ability to get my hands on the data. Meaning, there were multiple delays in actually receiving the data which pushed the project back. Not to mention, that the data that contained actual snow usage was missing which was very important to conclude the project. All in all, my experience was amazing and I am really thankful for the opportunity to be a part of this project.

My project only opens the door for further exploration and analysis. Again, my product only shows the distribution of routes against watershed areas. However, more in-depth analysis is required to better understand the salt usage in the county. I would recommend using ArcGis which would provide more sophisticated tools to analyze other features such as road width, elevation, and other GIS concepts. Another next step would be to continue tracking salt routes and their salt usage to gather sufficient data. This would allow the county to track how watershed health has changed over time and give insights into other areas of importance. Additionally, monitoring the amount of salt used is important to understanding which depots may be dropping excessive salt and establishing a connection as to whether the drivers are county drivers or hired contractors. Lastly, implementing route optimization techniques could help streamline salt distribution processes, potentially leading to cost savings and reduced environmental impact. Overall, these next steps would contribute to a more comprehensive understanding of salt usage in the county and inform strategies for more sustainable and efficient practices.

To conclude, I would like to acknowledge the individuals who provided support and made the completion of this project possible. My mentors, Jp Chamoun, Reza Zarif, Victoria Lewis, and Victoria Liu provided guidance and insights that led to the development of the data product. I am thankful for their support and encouragement throughout this semester. I would also like to thank my MC data professors who gave me the skills and tools to build such a product. They provided me with a strong foundation in data theory and coding application which I used to complete this project. Lastly, I would like to thank my fellow data science students who

provided words of encouragement and feedback to make the most of this semester. I am very proud to have worked with them and to see how much our skills have improved since the beginning of the program.