good argumentation of the problem. how do accidents, casualties and vehicles say something about road safety and quality how do you define this? How do you know that these attributes of the data correlate to economic faults? Good argumentation for the years. argumentation why visualization should be used for this could be improved. It is this a bit vague and general.

# Introduction:

Our tool, Zoom, is created for the Road and Safety Authority of the United Kingdom (R\&S) with the purpose of understand how the UK’s GDP and economical status effects the road quality and road safety for their citizen. We believe that it is vital for countries to create long lasting infrastructure which would not compromise on a road users safety. In lieu of such, we have chosen to focus on the question which is as follows:

To what extent did the 2007-2008 financial crisis effect the road safety and quality for the drivers in the UK?

To illustrate and explore the given topic, we will be looking at the dataset provided over the years 2005-2006, 2007-2008 and 2011-2012. The dataset is a comprehensive list of all accidents, casualties, and vehicles involved during the specified time periods. We have chosen the specified dates as the 2008 financial crisis took place through 2007 till 2008, through which the UK’s GDP was negatively affected as it dropped by 1.5 percent. Figure 1 [Crib, 2018] helps illustrate this point as the drop to the GDP has left long lasting effects for the country. As shown through figure 1, it took the UK till 7 to 8 years in order to regain where its GDP was present in 2007-2008. There were other additional concerns and developments because of the crisis which will be taken into consideration throughout the exploration.

The years 2005-2006 have been taken as a control group in order to comprehensively understand how the financial crisis affected the road safety of the UK. It is vital to understand if there was actually any change present through these years as it would allow us to discover trends and other testing methods in order to reach a final conclusion regarding the specified question. Additionally, the years 2011-2012 have been chosen to help illustrate and understand how further development of the roads in the UK were maintained and how such effects the road safety of the users. As stated before, 2011-2012 was before the period in which the UK regained its previous high GDP, and due to such, evaluation and understanding of the long-lasting concerns regarding road safety can be fully evaluated.

As previously stated, the purpose of Zoom is to aid in the exploration of any faults present with the current road infrastructure in the UK. We plan on specifically focusing on the economical domain aspect that affects the road safety and quality. Focusing on such is vitally important for long and short-term purposes. Discovering any major faults caused by economical failures of the UK will allow the R\&S to be able to better delegate funds to necessary infrastructure in times of economic prosperity. Through such measures, the road safety of road users will be preserved in times of economic downfall, allowing the government to focus their economic resources onto other vitally important aspects in order to regain their economical standing. Visualization is the correct way to answer the question that was specified. There are three main reasons why visualization is a beneficial way to answer the question and they are as follows. Firstly, through visualizations problems can be localized to localized sections of the UK through which further analysis and exploration could potentially be made. Secondly, through visualization, finding and exploring trends would also be possible over multiple attributes of the dataset. Lastly, visualizations allow R\&S to independently also check other aspects of the regarding trends and features found.

Data processing is limited. You could for example think about your time data, you could process it to display week or weekend days, or per day like Monday, this could be interesting while researching accidents. What do you do with missing data? Do not describe the attributes and their types, I assume you also have quantitative data or maybe categorical (such as discussed in the lectures?)

# Data and Task Analysis:

As previously stated, the overall goal of the visualization tool, Zoom, is to help R\&S inspect the data accident dataset over the course of 2005-2012. Through Zoom, R\&S plans on analyze the effects that the 2007-2008 financial crisis had on the road infrastructure of the as the financial crisis lasted for 5 fiscal quarters through which unemployment rose from 5.2\% to 7.8\%[David N.F. Bell, 2020].

Due to the nature of the purpose of the project, there are a considerable number of variables need to be accounted for in this project. We have chosen to focus on the following aspects as we believe they are essential in order to answer the purposed question: Severity of accidents, involvement of governmental agencies (Police, Ambulances, Firefighters, etc.) and accident location. We believe that through these basic attributes, it would allow R\&S to identify and select areas with outliers, trends, or extremes of any sort. The purpose of these attributes is to find anomalies. Other attributes will be shown for individual data points, but the above attributes will be compared and browsed through. Such attributes would be: number of accidents, severity of the accidents, governmental involvement etc. These attributes would allow for better of the situation on a case-to-case basis, allowing R\&S to analyze the conditions under which these anomalies occur. Analyzing these conditions would allow for better understanding of how infrastructure could be placed to prevent such accidents, supposing that the anomalies were present due to the lack of infrastructure.

Through the understanding of these attributes and preliminary research, we hypothesize that: “The financial crisis had a minimal effect on the road infrastructure in the UK”. We however would like to state that through preliminary evaluation of the dataset, there is a higher number of accidents through the years concerning the financial crisis. We do not believe that the road infrastructure itself was ill-maintained, but through other causes.

# Data and Task Abstraction

In order to answer the hypothesis, and the question as a result of that, we have also created specific low-level goals to further help decide the type of visualizations. Through the task abstraction, we see that we need to adhere to the following low-level goals:

* Compare Trends
* Browse Extremes
* Locate Outliers

As previously stated, we would primary like to identify trends, outliers and extremes through our visualization tool. After such, R\&S would be able to better identify correlation between such anomalies through the individual data of the incidents. We believe that R\&S should be able to independently determine if better infrastructure is required in order to decrease the number of accidents present in that area. We have chosen this approach as we believe not all accidents caused by poor infrastructure and further evaluation is needed for such case as there may be more effective ways that R\&S could implement in order to prevent such anomalies.

Zoom takes in tabular data from the provided dataset which contains spatial and geometrical data types, along with other aspects for a given data point. The dataset would be then reduced to the years that we have specified for the purpose of our tool. This process would involve only selecting years and attributes that we would like to represent on the visualizations. All attributes given in the main dataset are represent through some numerical value. A few of these attributes will be left as numerical values, but other values will be converted to their descriptions in order to better help the user understand their meaning through the visualizations.

For the purposes described in the task abstraction section, the data will be processed in order to achieve the task abstractions goals specified. The primary way the data will be processed in through grouping based on a certain set of attributes. Through grouping, the data will be able to highlight extremes easily while also showing trends around certain groups or areas. The data will otherwise have minor transformations as the visualizations can directly derive the necessary data.

# Current Visualization and Interaction Design:

There are two main indices that will be used in Zoom. Firstly, as we want to analyze and understand how the financial crisis had affected the road standards in the UK. Regarding this aspect, we believe that Zoom should be able to show ordinal and discrete time sections which the user can browse through in order to effectively focus on certain time sections to better analyze. Zoom will allow users to see the entire dataset of either 2005-2006, 2007-2008 or 2010-2011. We believe that it would be better to understand entire datasets and compare them accordingly. Being able to continuously browse the data does not fully allow the user to help isolate and identify incidents. Secondly, the users would also need another indication to evaluate regions or areas in which a greater number of incidents occur. For this purpose, there will also be a display a map of the UK through which the user would be able to select region and areas through which they can browse a set of datapoints through. The time index will appear on top of page, while the map will appear on the left side of the page and will take up a large majority of the screen. Other visualizations will be placed on the right side of the page in order to give more detailed information about the datapoints selected. There will be four main types of visualizations which will be used in our tool. They are as follows:

The first type of visualization, as previously described, would be the map of the UK with datapoints. When the map is fully zoomed out, the datapoints will be grouped together by region. The visualization will use the map as a mark as it will represent the data geographically. The color, size, and position will be used as channels as the visualization will use circles to convey information. The location of the circle on the map will identify where a group of accidents took place, the size of the circles will represent the number of cases that took place in that region, while the red opacity of the circle will help represent the severity of the group of accidents. Such a visualization is primarily used in order to discover features and locate outliers in the dataset. Such a representation is effective at locating outliers and discovering features as the user will be able to easily identify areas with larger circles, or less opaque circles. Such areas would be highlighted compared to other parts of the map, allowing the user to select these areas and investigate them further. We expect the map visualization to be similar to figure 2.

The second visualization technique implemented is pie charts. Pie charts are used in order to give statistics about the severity of the cases that are selected. As shown through the dataset, the severity ranges from no harm, to bruises, and gets to the extent of serious or fatal injuries. The marks will be the area of the circle each severity occupies, and the channels will be the colors to distinguish each severity. The pie charts allow users to highlight extremes as through selection of data, if a pie chart shows an incredibly high number of severe injuries then further evaluation might be considered.

The third visualization is a line plot which shows the number of accidents over time. Given the period of time that the user selected, the line plot will highlight the trend of the accidents that occurred throughout that time span. All peaks will be highlighted and the data will be localized based on the area selected through the map. The mark used will be a line that travels and connects the dots on the graph, while the channel itself would be the placement of data points. This visualization is used to compare trends.

The fourth visualization will be a horizontal stacked bar graphs which will compare the local and global values of three distinct attributes. Firstly, local refers to the area selected in the graph by the user while global refers all the data in the given time period. Secondly, the attributes to compare are the number of deaths, the number of accidents, and the number of government involvement vehicles present. Percentages will be used to show the extent of the local and global data being compared. The purpose of this graph is to compare distributions. Marks will volume of the bars filled which would be correspondent to the percentage, and channels will be the color of the bar as different percentages will display different colors.

Besides the main visualizations, we will also have a data panel which given the general overview about the data selected such as the number of accidents, the area which is selected, and the average severity of the accidents selected. The purpose of this panel is to help present features. We expect the compilation of these visualizations to look like figure 3. \\

Lastly, the color scheme of the visualization is going to be red, yellow, turquoise, and a light tone of purple. We want to use red, yellow and turquoise to help indicated the extremity of the features we plan to highlight while maintaining a safe display for users with color-blindness. Purple will be used to highlight any other features if necessary. The background of the visualization will be dark/navy blue as the chosen saturated colors will contrast well with the dark background, causing the users attention to be bound to specific highlighted features of the tool.