**Project Proposal – Team Cool**

1. **Project Title:** An Analysis of Covid-19 across USA.
2. **Team Members**
   1. Tyler Ward (tward90)
   2. Aaron Paul Lewis (lewisaaronpaul)
   3. Ross Davidson (rossdavidson47)
3. **Project Description/Outline**
   1. An analysis of Covid-19 cases and deaths by county in USA.
   2. Analysis of demographics, lockdown timing, weather, birth rate, transport patterns?
4. **Questions to Answer**
   1. How did the reported cases spread across the USA? What counties reported cases first? (adjust for population)
   2. What is the relationship between reported cases and reported deaths? e.g. did deaths start to rise at a consistent lag to cases? Per capita, by geography. Heat maps? Top 5 counties? By state?
   3. What is the impact of demographics by county on infection rates?
      1. Health Care shortage.
      2. Census Data - Poverty rates.
      3. Latitude.
      4. Lock-down.
      5. API scrape
      6. Heat Map
      7. Line/Scatter chart
      8. SPSS/Excel stats analysis
   4. What is the impact of lockdown timing on infection rates?
   5. (to explore if we have time) What is the impact on weather/latitude/longitude/temperate/Netflix/conception/ambalance response/911 calls/hospital availability/medical facilities/traffic patterns on infection/death rates?
5. **Datasets to be used**
   1. USA facts.org: <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>
      1. Covid\_county\_population
      2. Covid\_deaths\_usafacts.csv – reported deaths by day and county
      3. Covid\_confirmed\_usafacts.csv – reported cases by day and county
   2. Google Cloud: <https://console.cloud.google.com/marketplace/browse?filter=solution-type:dataset&filter=category:covid19>
   3. [WHO Dashboard](https://covid19.who.int/?gclid=CjwKCAjw8pH3BRAXEiwA1pvMsUvvfKq71lWhH-CrXgoqtUxu16vmVFHBmDks1krmjnEdtmbOy1-c8RoCgX8QAvD_BwE)
   4. [City-Data.com](http://www.city-data.com/)
   5. <https://covidtracking.com/api>
   6. [Financial Times](https://ig.ft.com/coronavirus-chart/)
   7. pre/post lockdown dates by county?
   8. Other demographics by county?
6. **Rough breakdown of tasks.**
   1. Data Exploration
      1. *Create a Jupyter Notebook describing the data exploration and cleanup process*
      2. *Use at least one API, with data pertinent to your primary research questions*
   2. Data Cleanup:
      1. *Use Pandas to clean and format your data set(s).*
   3. Data Analysis
      1. *Create a Jupyter Notebook illustrating the final data analysis*
   4. Data Display
      1. *Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)*
      2. *Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation*
   5. Summary and Conclusions
      1. *Create a write-up summarizing your major findings.*
      2. *This should include a heading for each "question" you asked of your data,*
      3. *Under each heading, a short description of what you found and any relevant plots.*
   6. Prepare Presentation
      1. [Presentation Guidelines](https://rice.bootcampcontent.com/Rice-Coding-Bootcamp/ru-hou-data-pt-04-2020-u-c/blob/master/07-Project-1/1/ProjectGuidelines/PresentationGuidelines.md)
      2. [Presentation Requirements](https://rice.bootcampcontent.com/Rice-Coding-Bootcamp/ru-hou-data-pt-04-2020-u-c/blob/master/07-Project-1/1/ProjectGuidelines/PresentationRequirements.md)
7. **Notes:** how do you cite sources?

**Project Plan**

**Saturday –**

Ross: data analysis and charts on existing data

Aaron: census API download, geo code and merge with case/deaths data

Tyler: merging healthcare data with case/deaths data

**Sunday –**

Aaron: explore census data, data analysis with case/deaths data

Tyler: family business

Ross: tbc

**Monday – Team meeting at 530pm CST**

Compare and consolidate analysis

**Tuesday – Class, focus on preparing presentation**

Weds – backup if needed

**Thursday – Presentation**

# #This code attempts to answer the following questions

#Question 1: Totals

#1a) How many cases and deaths have been reported in total since January?

#1b) How has the rate of change developed over time?

#1c) What observations can we make from this high-level analysis?

#Question 2: Saturation

#2a) How many counties reported cases and deaths each day?

#2b) ###Map of infected counties ###

#2c) What observations can be made from this analysis?

#Question 3: Effect on different States

#3a) What is the relationship between State population and Total Deaths?

#3b) Which states have the highest deaths/cases ratio?

#3c) Are any of these outliers (box/whisker)?

#3d) What is the distribution of Deaths/Case Rate?

#3e) Do the states with the highest death/case rate also have the highest deaths per capita? (A high death/case ratio could lead to fewer deaths per capita if cases die before they infect others)

#3f) Interesting observations from this analysis?

#Question 4: Concentration

#4a) Which States/Counties have been most affected?

#4b) Where are they?

#4c) Is there a correlation between Total Deaths and Lat?

#4d) Is there a correlation between Total Deaths and Lng?

#4e) Interesting observations from this analysis?

Here is the **schedule for Project Presentations for Thursday 25 Jun, 2020**  
We will have 4 slots as follows,  
**Slot 1 - 5:45 pm - 6:30 pm** (3 teams will present in this session. These teams will have the option of leaving the class early at 8:45 pm and not attend the QA session. ***Please let me know if you want to present in this time-slot and take advantage of the early finish***)  
**Slot 2 - 6:30 pm - 7:30 pm** (4 teams will present in this slot)  
**Break - 7:30 pm - 7:45 pm**  
**Slot 3 - 7:45 pm - 9:00 pm** ( 5 teams will present in this slot)  
**Slot 4 - 9:00 pm - 9:30 pm** (Q/A session for all teams. All teams present will take questions about their projects from students and staff. *There are no grading points for QA*)  
Note that only the 3 teams who are opting to present in Slot 1 have the option of skipping Q/A and leave class early at 8:45 pm. For everyone else, it is optional to attend Slot 1, but mandatory to be present from 6:30 pm to 9:30 pm.

**Maximum time limit for team presentations** - this will be **10 minutes or 14 minutes** depending on your team-size  
If you have a 3-member team, you have a max limit of 10 minutes  
If you have a 4-member team, you have a max limit of 14 minutes

* ***Only one person from every team will share screen throughout the team's presentation***
* ***All team members will take turns to present (approx. 3 minutes per person)***
* ***It is mandatory for every person in the team to present.***

*Try not to go over your code during your presentation. We will have the pleasure of looking at your code later*  
***Try to keep most of your presentation centered around the data story, inferences and the analysis you did to arrive at the story.****Also talk (briefly) about where and how you got your data and how you cleaned/processed it.* (edited)



