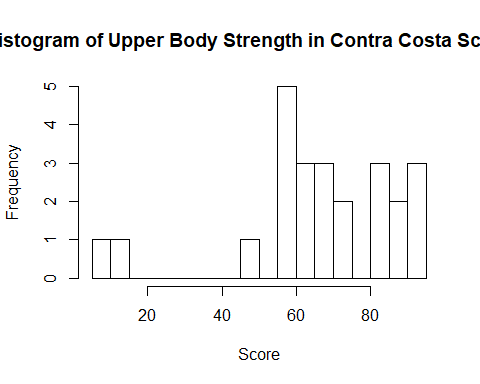
Data Science Team Vision

* Collaboratively develop a Data Science Team at John Sweet High School in Crockett, California
* Introduce interested students to the “hot” field of Data Science
* Decide together what web applications and/or other “products” to build based on interests and motivations of students
* Team mentor, Michael Samuel, will introduce the Team to a suite of tools based on the free and open source R ecosystem, including
  + R based software,
  + RStudio (graphical user interface)
  + tidyverse set of packages for “data wrangling”
  + R Shiny package for interactive visual display
  + tmap for mapping
  + and many others
* Michael will also introduce the use of GitHub as the repository for Team work, and to facilitate several other aspects of team collaboration
* The Team will then decide together what tools (e.g. software) to use moving forward
* Logistics
  + Some logistics will need to be determined based on interested students’ scheduled
  + Possible plan would be to meet one afternoon a week after school in the ??? room, and one afternoon a week “virtually”, with the mentor at home or at work, and the students at school or other convenient locations
  + With the students we may well can the name from “Team” to something they prefer. The idea of the “Team” is much less about competition with other schools (although maybe that’s an idea too…) but about working as a team together to learn, have fun, and maybe make some cool (or dope, or whatever a current word is…) and/or useful things together
* Initial Team GitHub Site is [here](https://github.com/mcSamuelDataSci/Data_Science_Team)
  + The document you are reading, in both HTML and .docs formats, can be found on this site, as well at the “R Markdown” file that created these documents (DS\_team\_introduction.Rmd)
  + All other current Team files can also be found here
* We will emphasize using “Open Data” and other data that is freely available in a machine readable format. The mentor will introduce these types of data focusing on community health and “Social Determinants of Health”, but after that, the Team could shift focus to any number of areas of interest like sports, music, technology or whatever
* Examples of Open Data related to California Schools that we might work with initially include:
  + [California Physical Fitness Test (PFT) program results:](https://www.cde.ca.gov/ta/tg/pf/pftresearch.asp)
  + [School metadata:](ttps://www.cde.ca.gov/ds/si/ds/pubschls.asp)
  + [Academic performance data:](https://www.cde.ca.gov/ta/ac/cm/)
* The California Physical Fitness Test (PFT) program results data are a very interesting and important data set. The data are broken down to the school, district and county levels, and include results of multiple physical tests and “body mass index”. Because of this, the data set can be used to explore many types of questions and to provide insightful and valuable analyses and visual displays.
* Part of the R code for working with these data, and some R code and resultant charts, is shown below.

test\_selected <- "Upper Body Strength"  
county\_selected <- "Contra Costa"   
selected\_info <- filter(fg\_districts\_info,   
 Line\_Text == test\_selected,   
 County == county\_selected,   
 NoStud5 != 0)  
  
hist(as.numeric(selected\_info$Perc5a), breaks=20,  
 main=paste("Histogram of",test\_selected,"in",county\_selected,"Schools"),  
 xlab = "Score"  
)



* R markdown resources:

<https://stackoverflow.com/questions/29787850/how-do-i-add-a-url-to-r-markdown/29788494>