

# Progress Report

Ryan Harvey and Max Narvaez

10/28/2020

## Updates to your proposal – what has changed in terms of websites scraped, variables acquired and generated, questions addressed, visualizations planned, and format of your final product

Not much has changed since our proposal. We have successfully scraped data from [The Blue Alliance](#). We have adjusted our scraping approach to just make it more efficient because we are working with some *very* large data sets. We pull the data once, then cache it to a .rds file that can be loaded later. Our variables acquired represent the ones in the table from the project proposal. The questions we want to address are still the same, but we may come up with more after working with the data more. As for the format of our final product, we will have a report with our methods, plots, and analysis. We will also produce presentation slides. If time allows, we will create a dashboard to better display our plots and results.

## An update on how your data scraping is coming

Like mentioned before, we have successfully scraped data for over 140000 matches between 2002 and 2019. We have been working on organizing this into data frames in optimal ways for us to produce plots.

Current tables are:

### **matches**

Contains one row per match played, with all data associated with that match in one row.

Variable Name	Description
year	Year that the competition took place
week	Week of the season that the event took place
event_key	Unique key for the event
event_type	Type of competition (i.e. championship, district, regional, etc.)
key	Unique key for a match of a specific event
match_key	Unique key for a match
comp_level	Level of competition (i.e. qualifying, eighths, quarters, semis, finals)
set_number	Number to identify different comp_level matchups
match_number	Number to identify which match during a competition
blue_score	Total score of blue alliance
red_score	Total score of red alliance
winning_alliance	Either “red” or “blue” or “tie” representing which alliance won
blue_alliance_(1-4)	Each of the team keys for the (up to) 4 blue alliance teams
red_alliance_(1-4)	Each of the team keys for the (up to) 4 red alliance teams

## scores

Same as **matches**, but each match now has two rows, one with the red alliance score and one with the blue alliance score.

Variable Name	Description
year	Year that the competition took place
week	Week of the season that the event took place
event_key	Unique key for the event
event_type	Type of competition (i.e. championship, district, regional, etc.)
key	Unique key for a match of a specific event
match_key	Unique key for a match
comp_level	Level of competition (i.e. qualifying, eighths, quarters, semis, finals)
set_number	Number to identify different comp_level matchups
match_number	Number to identify which match during a competition
winning_alliance	Either “red” or “blue” or “tie” representing which alliance won
blue_alliance_(1-4)	Each of the team keys for the (up to) 4 blue alliance teams
red_alliance_(1-4)	Each of the team keys for the (up to) 4 red alliance teams
alliance	The alliance whose score this row holds
score	The score achieved by the alliance

## teams

Same as **scores**, but now each team is in their own row, meaning there may be up to 8 rows associated with a match.

Variable Name	Description
year	Year that the competition took place
week	Week of the season that the event took place
event_key	Unique key for the event
event_type	Type of competition (i.e. championship, district, regional, etc.)
key	Unique key for a match of a specific event
match_key	Unique key for a match
comp_level	Level of competition (i.e. qualifying, eighths, quarters, semis, finals)
set_number	Number to identify different comp_level matchups
match_number	Number to identify which match during a competition
winning_alliance	Either “red” or “blue” or “tie” representing which alliance won
alliance	The alliance whose score this row holds
score	The score achieved by the alliance
position	The position in the alliance the team had (1-4)
team	The key for the team in this row

## Names of files where I can find your R code

We have all of our files in a git repo. Here is a link the the [GitHub](#). If you prefer we share these files with you another way, please let us know.