

Project Proposal

Lisa Fisher and Isaac Slagel

Website(s) you plan to scrape

We are interested in scraping data from Tour de France general classification winners. Websites include Wikipedia and a yet to be determined doping source. When we find a source for doping information we will hopefully merge that dataset with the wikipedia information, but that is yet to be found.

Variables you plan to acquire, including variables you plan to generate from your raw data

This data includes who the winner was, their home country, the year, their team, and a slew of information about the course. We plan to acquire if the athlete has been proven to be doping or not. We plan to generate replacement winners in cases of doping. We can also generate average speed.

Questions you plan to address

What individuals, teams, and countries have dominated the tour?

How prevalent is doping and how has it changed performance?

Do winners tend to also win points, mountain, or young rider classifications?

How many stages do winners generally win? How much of the race to they lead?

How has average mph changed over time?

Format of your final product and visualizations you envision

We plan to make a shiny app.

Our visualizations we can envision now are:

- a map of countries vs wins
- year vs mph
- bar plots which compare doped vs non-doped athletes
- many more