**ATP Tennis 2000-2023**

*Objectives*

* Determine who the real GOAT is

*Metrics*

* Number of grand slams/Number of all titles (Stacked Bar Chart)
* Highest ranking for longest period of time (Line chart – need to quantify number of weeks at number 1)
* Number of wins over losses
* How easily they won matches based on the score (Games won/games lost and sets won/sets lost)
* Number of wins over losses against other GOATs
* Number of wins over players with a higher ranking?
* Tournament variety of success (Packed Bubble Chart)
* Domination factor (early losses in tournaments, winning against high ranked players/losing against low ranked players, length of rankings)
* Any absences (injury, other reasons)

*Preprocessing*

* Add weight to different series of tournaments (Grand Slams have higher weight, masters 1000’s, etc.)
  + International = ATP 250
  + International Gold = ATP 500
  + Masters = Masters 1000
  + Masters Cup = Grand Slam
* Split score into Set\_1, Set\_2, Set\_3, Set\_4, Set\_5 and also create games\_won and games\_lost.
* China Open 2013 year was as 2012 so manually updated the correct records to be 2013.

*Visualization*

* Number of total titles including grand slams and surface of court
* Ranking over time – analyzing any dips and long periods at 1
* Frequency count tables of total career record, record against other goats, record against higher ranked players, record against lower ranked players, early losses in tournaments
* Box plot of games won over games lost