Grand Slam Heroes

Ganesh Viswanathan and Roma Dutta

2019-12-09

Chapter 1 Grand Slam Heroes

"The Open Era is the current era of professional tennis. It began in 1968 when the Grand Slam tournaments allowed professional players to compete with amateurs, ending the division that had persisted since the dawn of the sport in the 19th century." - Wikipedia

Github Link: https://github.com/ganesh2512/finalProject
Rstudio Cloud Link: https://rstudio.cloud/project/704614
Bookdown Link: https://bookdown.org/rdutta4/bookdown-grandslam/
ShinyAppsIOLink: https://ganesh-viswanathan.shinyapps.io/finalProjectShiny/
Data source Link: https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-04-09

Chapter 2 Introduction

"The Grand Slam tournaments, also called majors, are the four most important annual tennis events. They offer the most ranking points, prize money, public and media attention, the greatest strength and size of field, and greater number of"best of" sets for men. The Grand Slam itinerary consists of the Australian Open in mid January, the French Open around late May through early June, Wimbledon in June-July, and the US Open in August-September. Each tournament is played over a two-week period. The Australian and United States tournaments are played on hard courts, the French on clay, and Wimbledon on grass." - Wikipedia

One of the significant benefits of Data visualizations is making big or small data easier for the human brain to understand, that makes it easier to detect patterns, trends, and outliers in groups of data. Good data visualizations always place meaning into complicated datasets so that their message is clear and concise.

Our application leverages some of the concrete analysis examples and provide efficient, effective, functional, and convenient model for users.

Domain Problem Characterization:

The purpose of our project is to find out the best player in any era of tennis by birth year, classify them according to courts further drill down into gender and by year of tournament. We then next look under the hood to ensure the visual works as designed with good performance and quality.

```
knitr::include_graphics('images/Capture_awesome.PNG')
```
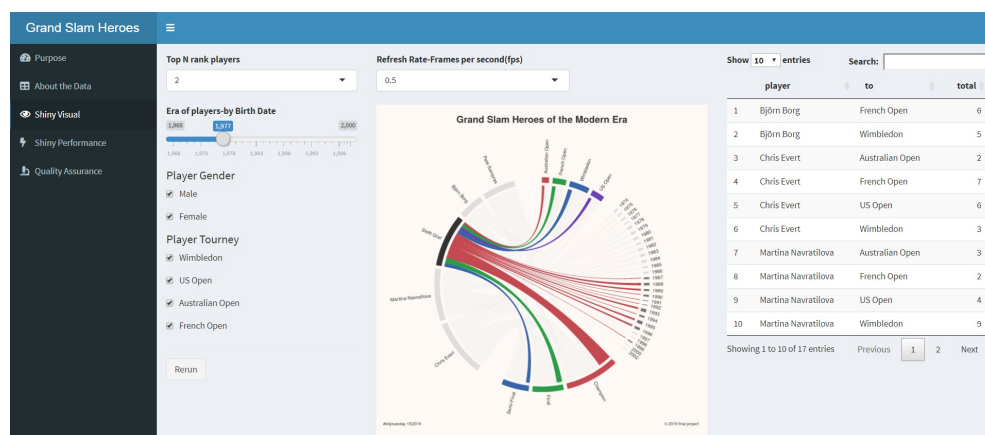
Chapter 3 Data/Operation Abstraction Design

Here we pull the data from source systems first as separate methods and break them down into simpler details which can be readily analyzed. The data operations are effectively abstracted out as R functions and reactive methods of shiny, hence increasing the efficiency of the visual.

The data has been taken from Dr. Torsten Sprenger's tidytusday post on Apr 9, 2019. We have taken help from tidyverse ecosystem to make meaningful charts with ggplot2, tidyr and dplyr in order to summarize and arrange data

```
player_dob <- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   grand_slam = col_character(),
##   date_of_birth = col_date(format = ""),
##   date_of_first_title = col_date(format = ""),
##   age = col_double()
## )
```

```
grand_slams <- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday
```

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   grand_slam = col_character(),
##   name = col_character(),
##   rolling_win_count = col_double(),
##   tournament_date = col_date(format = ""),
##   gender = col_character()
## )
```

```
grand_slam_timeline <- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/ti
```

```
## Parsed with column specification:
## cols(
##   player = col_character(),
##   year = col_double(),
##   tournament = col_character(),
##   outcome = col_character(),
##   gender = col_character()
## )
```

Chapter 4   Design Approach

4.1   Encoding/Interaction Design:

– User friendly Shiny dashboard is very convenient to show visual representations interactively. It addresses more specific questions seamlessly. As redundant visual representations may negatively affect its usefulness, we have aimed to provide a concise visualization with deep understanding.

– The application has information of the four grandslam tounaments starting from year 1968. Hence, one of the key interactions to be implemented for the user was the ability to select a Tournament and Year range.

– In this application, we have few inputs fields, output visuals and methods - reactive and non reactive methods which are modular and layered to create a effective visual.

```
knitr::include_graphics('images/Capture_interaction.png')
```
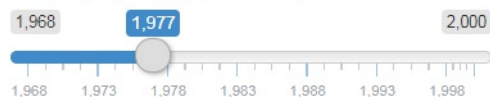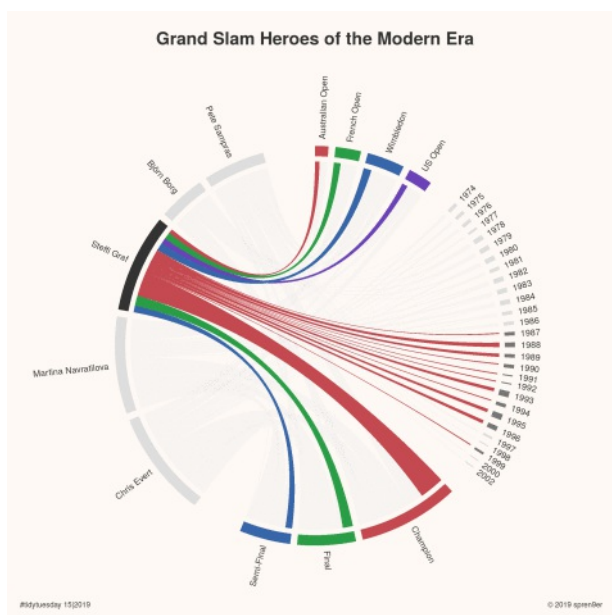


```
knitr::include_graphics('images/download_image.gif')
```

4.2 Algorithmic Design:

– Design is the branch of discrete mathematics and computer science that deals with the research, development and implementation of sequential and asynchronous algorithms. … Algorithms are used in every field that deals with values that can be quantified and many fields that deal with values that cannot.

– The prime objective behind any design is to create an algorithm to carry out the visual encoding and interactive design effectively. The performance of the system is significant component of the accessibility and usability. We have coded and designed the system considering the optimized performance of the application. The variables and dataframe which may slow down the application has created as a pre-processing step. Additionally, we have also given reproducibility utmost importance.

In a nutshell, we have taken the following into design consideration -

1. Reactive Method:

We incorporated the process of calling nested reactive methods instead of monolithic modules to improve the speed of processing and make sure only less amount of code need to execute for small changes in input like layers of an onion peel.
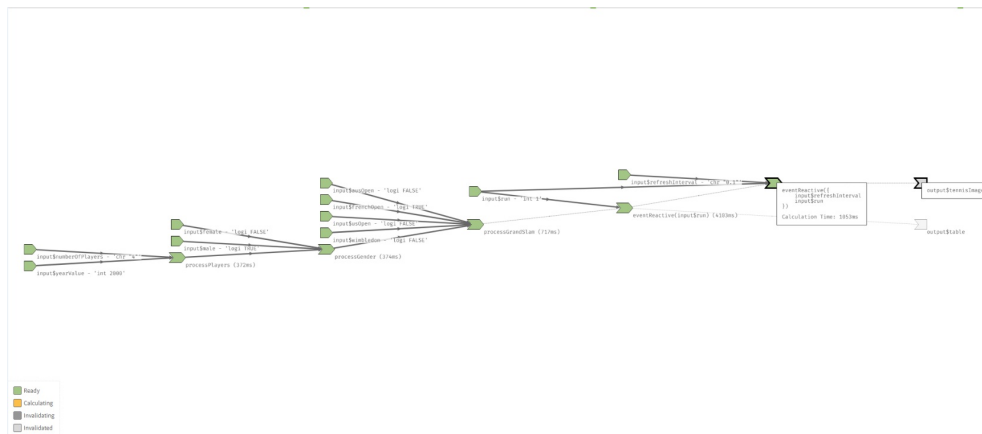
2. Caching Mechanism:

Caching of data for repeated calculations

3. Modularity:

To maintain the code in an efficient fashion

```
knitr::include_graphics('images/Capture_profiling.png')
```

Chapter 5 User Evaluation

– User experience evaluation (UXE) or user experience assessment (UXA) refers to a collection of methods, skills and tools utilized to uncover how a person perceives a system (product, service, non-commercial item, or a combination of them) before, during and after interacting with it.

– We would like to reach out to the Tennis enthusiast in the form of survey using social media. This will play as an essential part for any future work to improve the quality of the application.

– The primary goal of the dashboard is to explain how the user may benefit from our system. As the User experience is pretty significant for success of any application, we would like to reach out to the individuals irrespective of their knowledge on GrandSlam. Keeping this thing is mind, the APP should be simple yet useful to draw insight. The evaluation ideally should consider whether the product meets the specific requirement of the user but that alone is not sufficient. The product must be effective, efficient and server purpose before it will be adopted by the final user.

– Here we have analyzed the usability of this design and made enhancements in areas which can be useful
1. The timing of refresh rate of the visual
2. Use of the rerun button
3. Automation testing setup and execution to ensure the app is working after every deployment

```
knitr::include_graphics('images/Capture_refreshrate.png')
```



```
knitr::include_graphics('images/Capture_automatedtest.png')
```
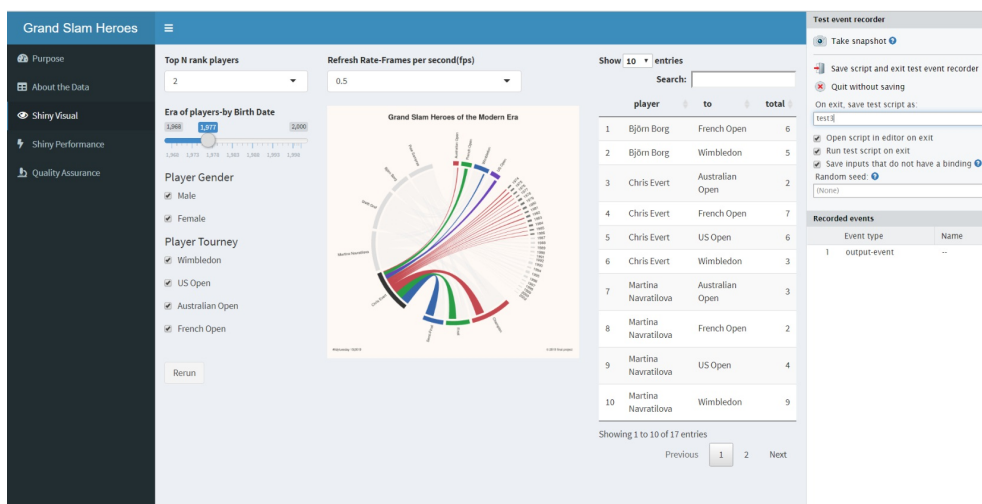
Chapter 6 Future Work

1. Improvement more on performance - Multi threading
2. Testing with complex inputs
3. Performance/Load testing
4. pre processing inputs to cache most of needed values to increase speed

5.. To get the tournament performance at age rather than simply across time. For this, we need to join the Date of Birth dataset with the grandslam dataset.

```
knitr::include_graphics('images/capture.png')
```
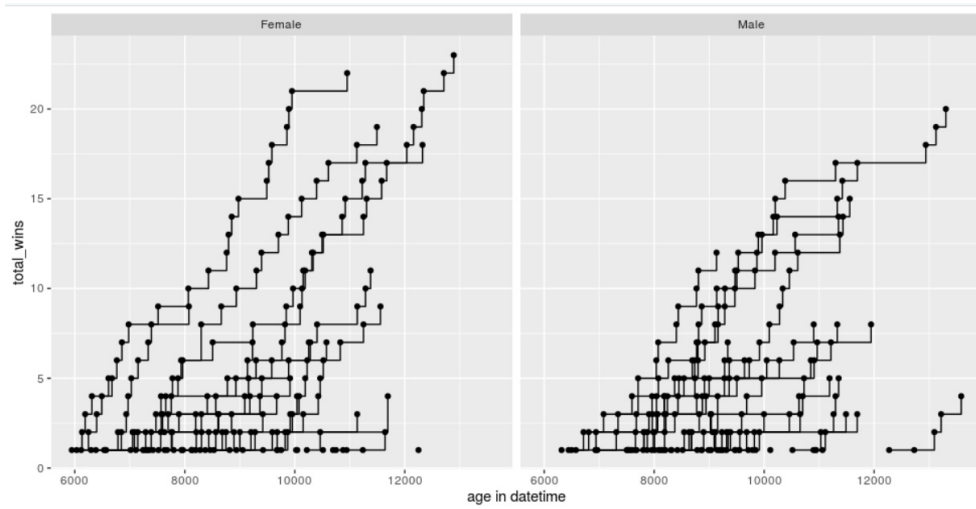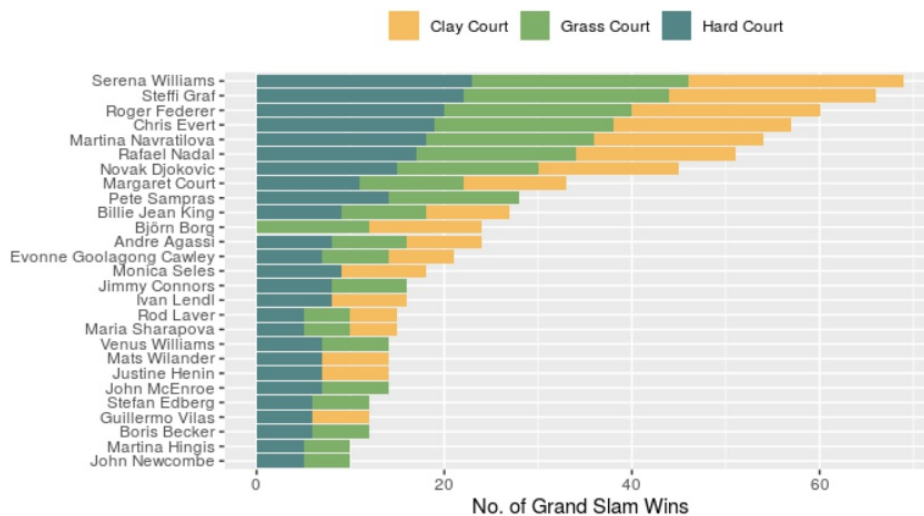

6. Come up with predictive model to predict win probability depending on the play court type

```
knitr::include_graphics('images/Capture_courttype.png')
```

References

Wikipedia: https://en.wikipedia.org/wiki/Grand_Slam_(tennis) Data source -Tidy Tuesday

https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-04-09

Dr. Torsten Sprenger's tidytusday post on Apr 9:

https://github.com/spren9er/tidytuesday

https://github.com/spren9er/tidytuesday/blob/master/tidytuesday_201915_tennis_grand_slams.r