Process Book

FIS Alpine Ski World Cup Viewer

Beuchat, Bastien Jollès, Eric Mamie, Robin

COM-480: Data Visualization (Spring 2020)

École Polytechnique Fédérale de Lausanne, Switzerland

2020–05–28

EPFL

Contents

1 Process breakdown **2**
- 1.1 Project's twilight . . . 2
  - 1.1.1 Data scraping . . . 2
  - 1.1.2 First ideas . . . 3
- 1.2 Main development . . . 4
  - 1.2.1 Athlete profile . . . 5
  - 1.2.2 Ranking evolution . . . 5
  - 1.2.3 Race map . . . 5
  - 1.2.4 Similarity graph . . . 5

2 Challenges and design decisions **6**

3 Peer assessment **7**
- 3.1 Bastien Beuchat . . . 7
- 3.2 Eric Jollès . . . 7
- 3.3 Robin Mamie . . . 7

4 Discussion **7**

Introduction

The FIS (*fédération internationale de ski*) Alpine Ski World Cup gathers sport fans across the world since 1966. Famous ski resorts from Argentina and Switzerland to New Zealand and China have hosted these events for more than 50 years. Although information is readily available on the official FIS website[1] and other third-party, fan-made hubs,[2;4] these resources fail to provide modern, dynamic and insightful visualizations.

To help providing new investigation angles to followers of the sport, we implement several original visualizations. A bar chart race shows the evolution of the World Cup overall rankings throughout each season since the beginning. The location of ski races are shown on an interactive map. When selected, the race results are shown on another panel. Skiers can easily be searched in our own database, while graphs retracing their performance season by season are displayed. Last but not least, a graph shows how professional skiers are *linked* via their specialty, showing that clusters of either versatile, technical or speed skiers exist in each season.

Our work succeeds in providing new insights by simply using simple, existing information on ski races and athletes. The platform is easily expandable and can be updated in a matter of minutes thanks to our scraping pipeline.

1 Process breakdown

1.1 Project's twilight

Data visualization takes up a big role in sports media. It is a very effective means of conveying information that once were only tables and textual statistics. Modern times have seen an explosion of such visualizations,[3] but this evolution seems to have avoided the ski world. The three authors being fans of alpine skiing – and still in mourning after the premature end of the 2019-2020 season –, it is only natural to have ventured in this area to contribute to the growing field.

This is why this projects explores the results of all FIS Alpine Ski World Cup results, from its inception in 1966 until today, after the shortened 2019-2020 season.

1.1.1 Data scraping

Getting reliable data on the World Cup results is our first goal. The FIS website[1] offers all statistics needed to reach our objectives, but the absence of an API, downloadable datasets or easily *scrapable* web pages forced us to find an alternative. Ski-DB[2] offers a simpler web layout which was much easier to scrape. The data has been observed and tested to be very reliable, if not identical to the official source.

Run-specific information				
Season	Date	Venue	Country	Event
Athlete-specific information				
Rank	Name	Country	1st run time	2nd run time
Total time	Time diff. to 1st rank		Ski brand	Ski-DB ID

Table 1: Information gathered for each World Cup result since 1966 (117,839 data points)

Name	FIS ID	Birthdate	Country
FIS competitor ID		Profile picture	Club

Table 2: Information gathered for each skier since 1966 (3,139 athletes)

Table 1 shows the information we have gathered for each skier's result, while table 2 is more specific for each skier. The information on the country is not redundant, since skiers may have changed nationality during the course of their careers. The difference between the *FIS ID* and the *FIS competitor ID* was essential during the scraping for obscure reasons only the FIS knows; it is only kept in case of *emergency*. These two datasets, once combined, may be used to build very strong visualizations.

Figures 1 and 2 show our early analysis efforts. These demonstrate that there are approximately between 60 and 90 events per season since the start of the World Cup, and that types of events have also evolved to leave more room to more technical disciplines, like slalom and giant slalom.

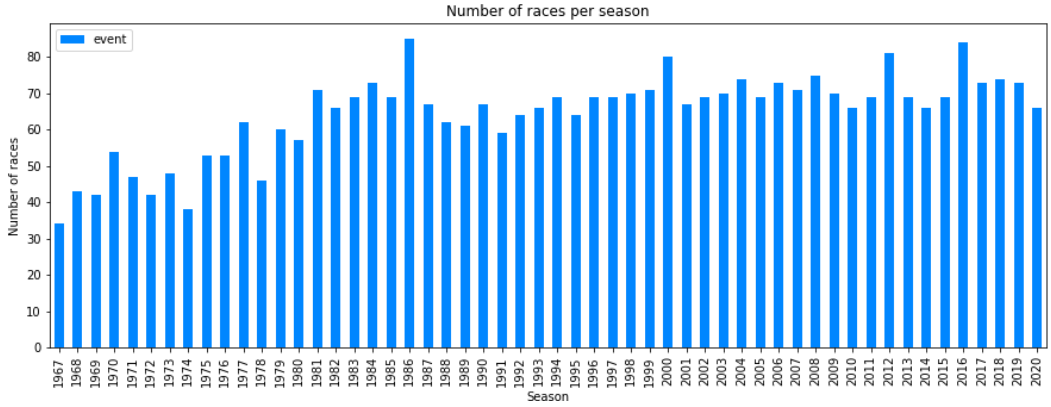

Figure 1: Number of races per season, period 1967-2020

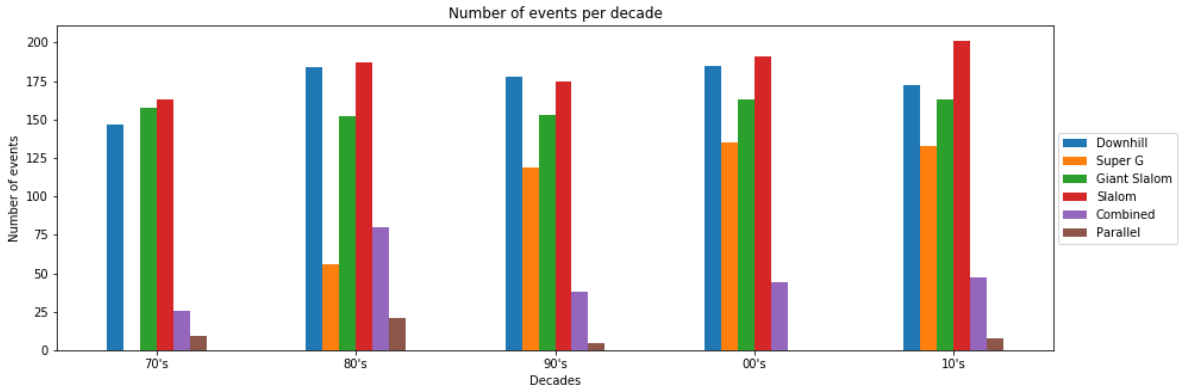

Figure 2: Number of events per decade, period 1970-2019

1.1.2 First ideas

Knowing the information we have at our disposal, we may now outline what exactly to visualize on a higher level.

Having such a long and detailed span of information allows us to show the results of absolutely every ski race ever. This is not – and should not – be the central point of our concept since explored sources already do this, but it still has to be present in one way or another. Another way of displaying these events is by using a map, dynamically showing the location of all the venues of a given season.

Having every World Cup result allows the creation of potentially interesting visualizations. Past rankings are nowadays sentenced to be stored as boring text tables, reflecting absolutely nothing about the suspense and tension experienced at the time. Dynamizing these rankings allows veteran fans to relive these moments as they happened at the time. Race after race, event after event, month after month, the rankings will evolve and give the context to each race; for example, "why did this slalom specialist participate in a downhill event? Oh right, it is because he is one of the main contenders of the main classification before he injured himself." Such a scenario is not shown by a static, end-of-season ranking table.

Finally, there is a possibility of linking this gold mine of information in a completely novel way. How do athlete compare among themselves, i.e. which skiers participate in the most varied type of events, which only stick to their specialty, etc. This way of

comparing and linking the athletes highlights the main rivalries of the time.

Figure 3 demonstrates an early idea for our website. The structure remained the same throughout development, since it proved to be an efficient way of displaying everything.

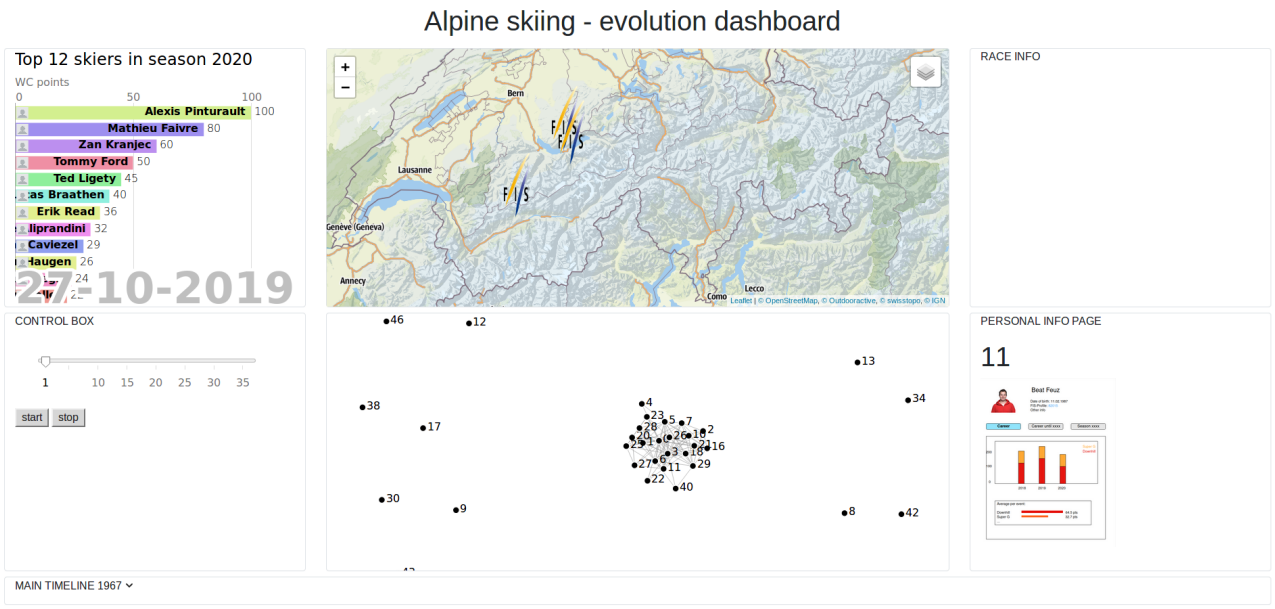

Figure 3: Early prototype of our website

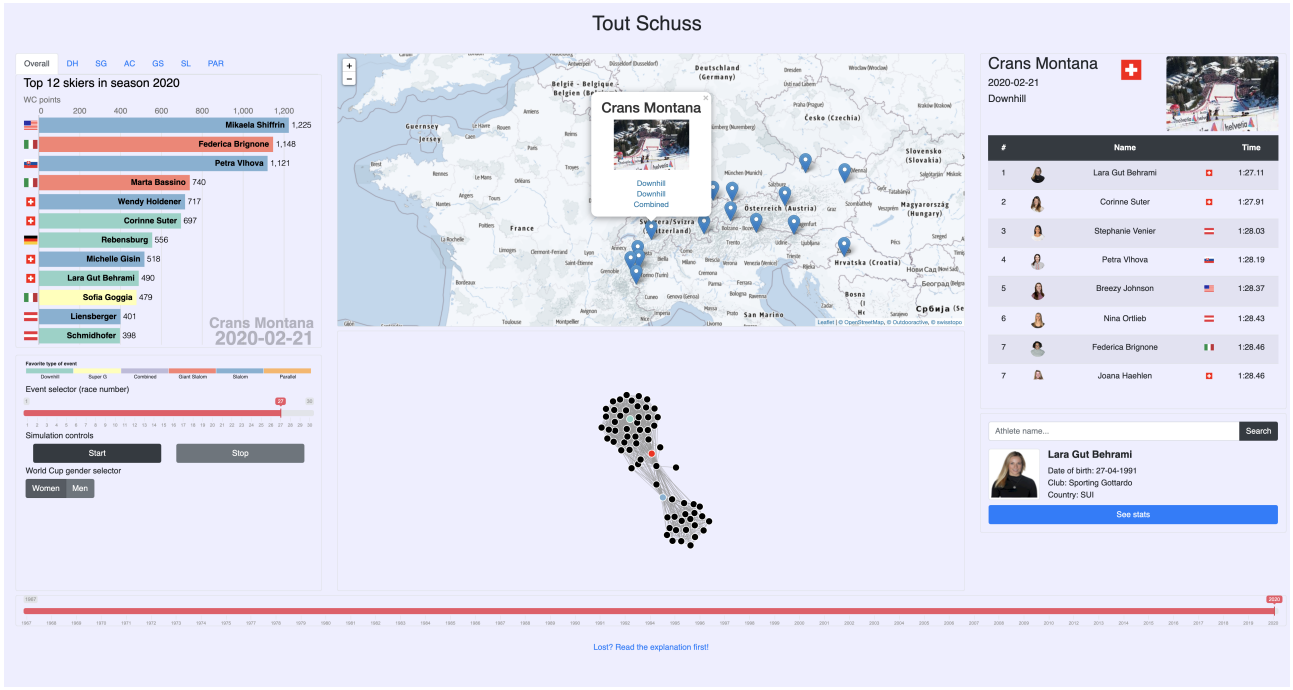

Figure 4: Final version of our website

1.2 Main development

On a high level, the user can select any season (1967-2020) and the desired gender of the competition they wish to explore – men's or women's. Figure 4 shows our final result, for reference.

1.2.1 Athlete profile

All skiers have a personal information page displaying their name, birth date, profile picture, club and country. This page can be loaded by either selecting a skier in any other visualization, i.e. the graph similarity, and the bar chart race or normal race rankings. A search bar with autocompletion is also available. There are also different visualizations available for each skier under three different modes, either showing their whole career, their career until the *selected season*, or their results of the season taking place on the *selected season*. Our data is dependent on what exists on the official FIS-website, so it is entirely possible that some information is absent or wrong for some athletes.

These visualizations are composed of a global point tally counting all the World Cup points collected by the skier and specialized point tallies, showing the mean number of points collected during specific events (slalom, downhill, etc.).

1.2.2 Ranking evolution

In order to show the evolution of the general classification rankings in any given WC season, we are using a *bar chart race*. Our graph is based on this visualization. A *bar chart race* is the animation of a bar chart over, generally, a given time. This allows to easily view the evolution of any dataset throughout its existence. We have added – compared to the base visualization – a slider allowing the user to move between race results of a given season. The flags of each athlete's country are added, and the color of any bar is dependent on the event specialty of each skier. On top of the general classification, there are also bar chart races for each discipline. A click on any bar shows the skier's profile in the aforementioned section.

1.2.3 Race map

The central world map shows all the events of a given season. All events taking place in any location are shown as a popover when clicking on the different locations, and further race details are shown in the top-right panel after selecting them. The map is also linked to the bar chart race; it is also animated, i.e. the map shows the event currently *updating* the rankings.

1.2.4 Similarity graph

A *similarity* graph shows the links between skiers, i.e. which skiers compete together in similar events. A graph corresponds to the *topology* of a particular season, showing which skiers are completely specialized – e.g. only participate in slalom events – and which are truly versatile. Being able to browse through the seasons and compare the evolution of the topology is insightful and gives a lot of information for ski fans.

Each node of the graph corresponds to a skier and a link is added between two skiers if they have raced against each other at least **K** times during the season. **K** is a threshold allowing us to drop less relevant links. This threshold has to be adapted through the years, because there are far more races and skiers in 2020 compared to the first seasons – we will discuss this detail in the *Challenges* section. The nodes are clickable and load skier information directly in the profile page. Colors highlight important nodes, for example the winner of a final ranking of the selected season.

2 Challenges and design decisions

With so many visualizations, finding a way to display them all in a meaningful way is quite challenging. We opted for a one-page control panel displaying all our widgets, working best for Full-HD monitors (1920 by 1080 pixels or more). Having a more responsive implementation would have cost us a lot of time, and would have been outside of the scope of this project. We find that all components are closely linked together, and displaying them so eases the task of the user when they want to find insightful links out of visualizations.

The similarity graph got quite computationally heavy for later years – there are more athletes and more runs to deal with. To resolve this, we simply set a stricter threshold for similar skiers and eliminated all that were not connected to at least one other skier. In-depth analysis – see figure 5 – has been made to choose the most interesting and the least demanding threshold. There exists a fine balance between responsiveness and data shown.

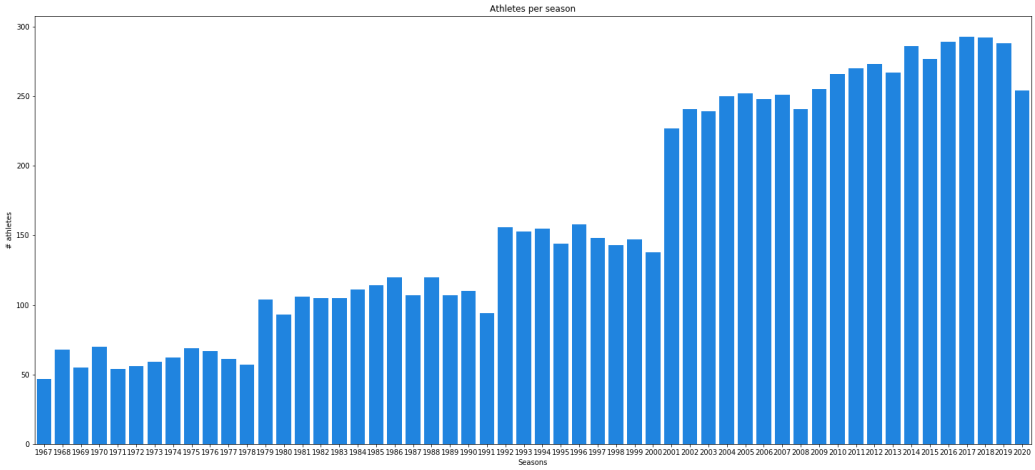

Figure 5: Analysis of the required thresholds for the similarity graph

Because of the way the FIS-website is designed, some skiers do not have a profile picture or personal information, although it is still stored on the website – we simply missed them during scraping for diverse reasons. We simply use a placeholder image for such cases.

We use Open Sans throughout our project, be it in this very process book, on the website, or in our screencast. Colors of events have been harmonized throughout the whole project. For example, a node of color in the similarity graph signifies that the skier has won a World Cup ranking during the season. We use external images for the skier profiles, so we had to make sure that they had the correct size when they appear in our visualizations.

The website is based on a grid layout, built with Bootstrap. This tool allowed us to quickly merge and harmonize all of our work. It provides useful user interface (UI) components for a uniform style, based on the best practices of the web. We also used it to code the modal of the profile statistics. As a technical detail, we only used one `iframe`, to display the screencast directly from YouTube.

All of the data displayed on our website is precomputed and present in their own json files. This is of course for performance reasons, as otherwise the site would take ages to load each time we switch the season or the gender. This increases the amount of transferred data, but only files used for the current visualization are loaded. Thanks to the browser's cache, the user can easily reload the same visualizations and still have that data locally.

We use the following libraries, imported via content delivery networks (CDN), in our project:

- `d3.js` for the graphs
- `Bootstrap` for the style and layout
- `Popper`, as a dependency of bootstrap
- `ion.rangeslider` for the two sliders (which has a slight bug in the tick alignemnt unfortunately)
- `Leaflet` for the map
- `JQuery` for the DOM manipulation

3 Peer assessment

3.1 Bastien Beuchat

Bastien is the certified web designer of our group. On top of doing a good chunk of data analysis and his superb similarity graph, linking athletes that perform in similar events, he has also taken on the gargantuan task of linking each component together, integrating them on one web-page and make everything look nice.

3.2 Eric Jollès

Eric has also a substantial experience with Javascript, being well-versed in Leaflet thanks to his past semester projects and the applied data analysis class. He has coded the map and the static race classifications, alongside looking for images for each of the more than 200 events scattered across the years and the world. He has also developed the bar chart race.

3.3 Robin Mamie

Robin has taken on the job of looking for, scraping, cleaning and organizing all the datasets used throughout our project. He has also programmed the athlete profiles, allowing the user to select any athlete from any visualization, or directly looking for them via text, and display their data and statistics. He has also written and organized the process book, and done the screencast of the project.

4 Discussion

Our *evolution board* fully achieves our initial goal, which is to create original and diverse alpine skiing visualizations. These allow any user to explore the whole history of the World Cup in an easy and intuitive way.

There are still other ways available to explore the data. Most notably, an athlete com-

parison panel could be interesting. Creating parallels between two champions of two different eras would provide additional information about the evolution of the sport.

The properties of the graph can also be exploited to summarize the style of the skiers throughout their careers, by not only evaluating their specialty, but also how their concurrents evolved at the time.

References

[1] Fédération internationale de ski. The Official FIS Athlete Database. https://www.fis-ski.com/DB/general/biographies.html.

[2] Mateo Pacor. SKI-DB™. https://ski-db.com/.

[3] C. Perin, R. Vuillemot, C. Stolper, J. Stasko, J. Wood, and S. Carpendale. State of the Art of Sports Data Visualization. *Computer Graphics Forum*, 37(3):663–686, 2018. https://openaccess.city.ac.uk/id/eprint/19857/.

[4] S. Pomerantz. Ski Reference. http://ski-reference.com/.