Process
Book

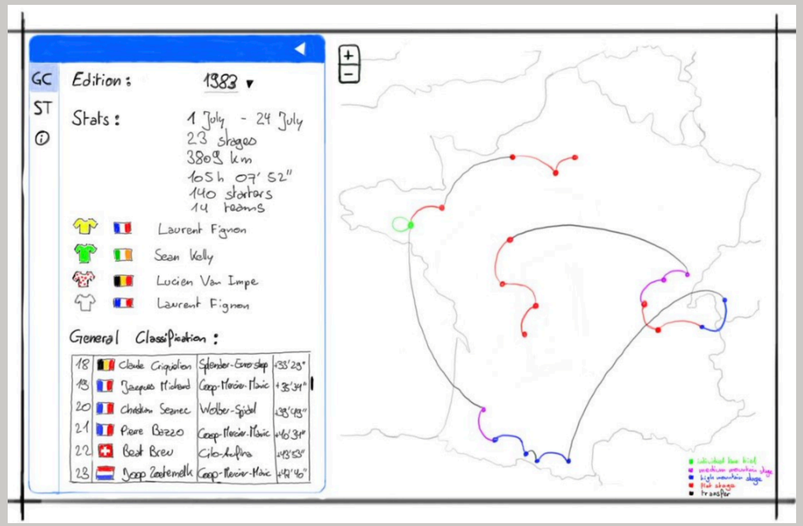
Tour de France

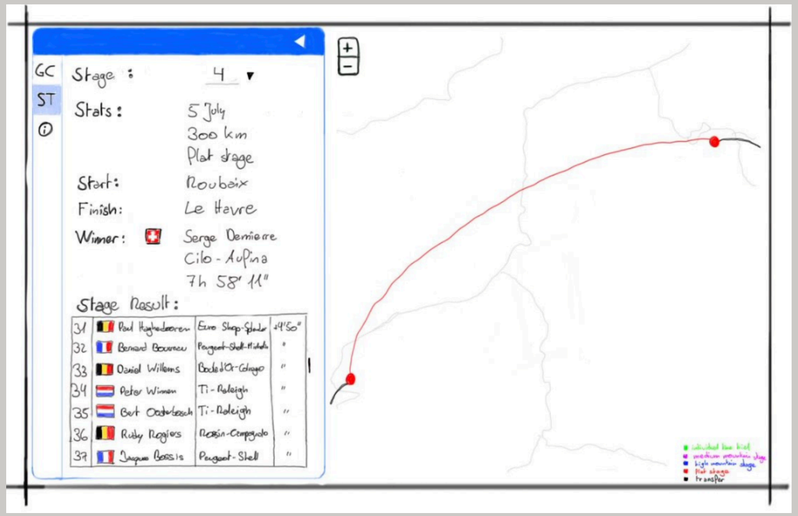
Jonas Konrad

Maxime Jan

Nana Tian

Concept





Motivation

Our aim was to build a visualization covering multiple types of data in order to try and implement more than one aspect of the theoretical knowledge gained in this class. The Tour de France, a famous yearly bike race, is a good fit for this.

Indeed, since its first edition was held more than 100 years ago, there is a clear evolution in time. With our visualization, we aim to make this legacy apparent and showcase the race's evolution through time. This evolution concerns all the aspects of the race, from the roads taken, the rules, to the number of participants.
By its nature, the Tour de France has a strong spatial dimension that can be clearly presented. Indeed, as the name suggests, the race is in essence a round trip of France, even if this is less apparent in the last few decades.

Finally, it also has a more statistical and numerical nature through the large amount of race results available. Trying to present this data coherently and without overwhelming the user is challenging. There is a lot that can be displayed, but we believe that less might be more in our case. Since not every user is familiar with the sporting achievements involved in such an event, trying to present data that might be too precise and of more interest to cycling fans might throw them off.
Nonetheless, the human nature of the race is still very important and can be highlighted. Indeed, every racer has a name and every performance is noteworthy.

Structure of the Visualization

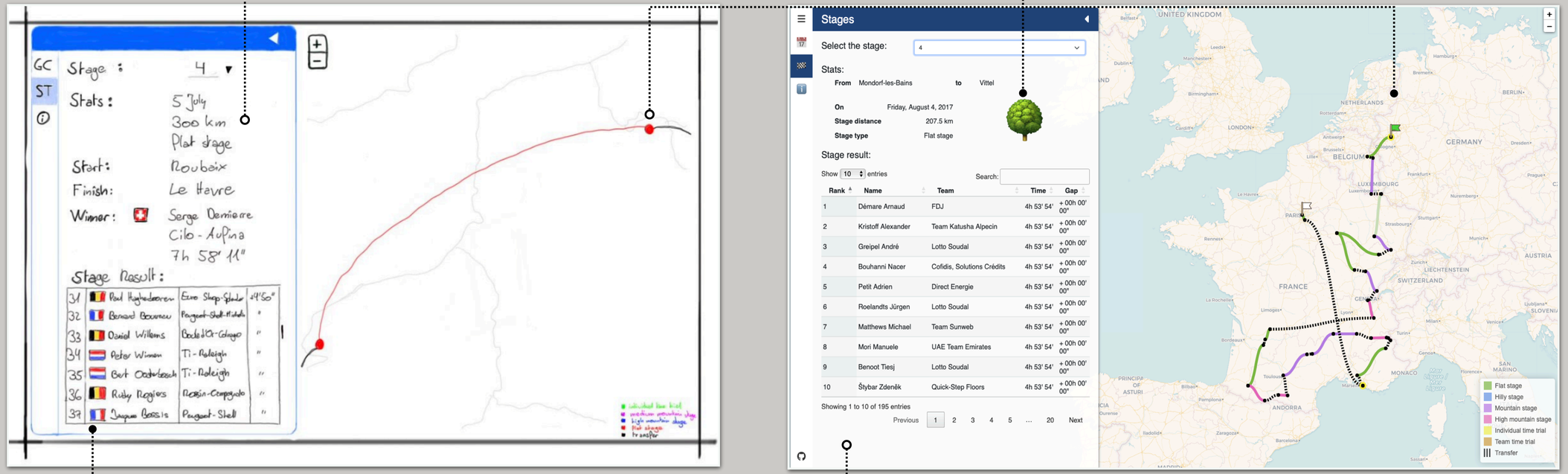
The obvious way to present geographical data is on a map. Our data contains the start and end points of every stage for every edition. Sadly, there exists no readily available data of the exact path each stage took. We hence decided to display on the map, for one edition at a time, the approximate path of that edition. To that end, we highlight the start and end point, and draw a line between them. Stages can be very different (distance, rules, amount of climbing, ...) and it is also worth showing that. To that end we coloured the stage according to their type.

To display the more numerical side of the data, we chose to use a sidebar. This allows the user to have both aspects displayed on screen simultaneously, highlighting the relationship between them, whilst at the same time giving them the choice to focus on any one of them. We use the sidebar to display the results of the edition/stage and some key statistics about it.

Finally, since we aim to accommodate for the casual user, we need a place where they can find the most important background knowledge to get the most out of the visualization. An information tab with textual explanations seemed to be the most efficient choice.This reasoning lead us to the sketches of the prototyping phase as they are presented in this document.

Changes from Design Prototype

Since Milestone 2. We were able to present all the expected information in the tabs, which includes a general review of status in both edition and stages as in the concept sketches. In detail, we also added symbolic icons to illustrate the stage type. For example, we use “⛰️ ” to give a vivid illustration of the high mountain stage. Likewise, we used ‘🌳 ’ to better picture a flat stage. More importantly, we completed the information tag to provide an overview of the tour de France. Also, on the map, we added a checkered flag in the map to indicate the finish of the edition and a green flag to indicate the start of the edition. In the sketches we show the nationality of each rider using his country's flag. In the end, our visualization does not contain this information because it is actually not available in our dataset.



Stage Tab

Stages

Select the stage: 8

Stats:

From Pau **to** Bagnères-de-Luchon

On Tuesday, August 9, 2016

Stage distance 184.0 km

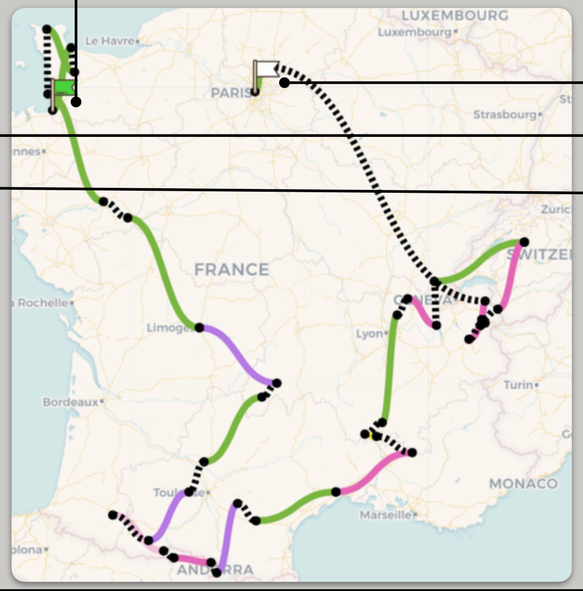
Stage type High mountain stage

Dropdown menus that allow user to choose the stage of interest

The stages status will update itself when choosing a certain stage

Statistics

Map of stages



The start of an edition

The end of an edition (Note: This should be a checkered flag due to the browser used it's white now.

Stage result:

Show 10 entries Search:

Rank	Name	Team	Time	Gap
1	Froome Chris	Team Sky	4h 57' 33'	+ 00h 00' 00"
2	Martin Dan	Etixx - Quick Step	4h 57' 46'	+ 00h 00' 13'
3	Rodríguez Joaquim	Team Katusha	4h 57' 46'	+ 00h 00' 13'
4	Bardet Romain	AG2R La Mondiale	4h 57' 46'	+ 00h 00' 13'
5	Kreuziger Roman	Tinkoff	4h 57' 46'	+ 00h 00' 13'
6	Aru Fabio	Astana Pro Team	4h 57' 46'	+ 00h 00' 13'
7	Yates Adam	ORICA-BikeExchange	4h 57' 46'	+ 00h 00' 13'
8	Valverde Alejandro	Movistar Team	4h 57' 46'	+ 00h 00' 13'
9	Mollema Bauke	Trek - Segafredo	4h 57' 46'	+ 00h 00' 13'
10	Porte Richie	BMC Racing Team	4h 57' 46'	+ 00h 00' 13'

Showing 1 to 10 of 198 entries

Previous 1 2 3 4 5 … 20 Next

The drop down menu allows users to select how many entries they want to be displayed in the current table

The search bar can allow search with rider names and/or teams

This table provides detailed information about the ranking of riders, their team, time to finish a stage and time gap from the winner

Edition Tab

Edition

Choose the year: 2016

Stats:

From Tuesday, August 2, 2016 **to** Wednesday, August 24, 2016

Number of stages	21
Total distance	2848 km
Number of teams	23
Total starters	199

Chris Froome

Peter Sagan

Rafał Majka

Adam Yates

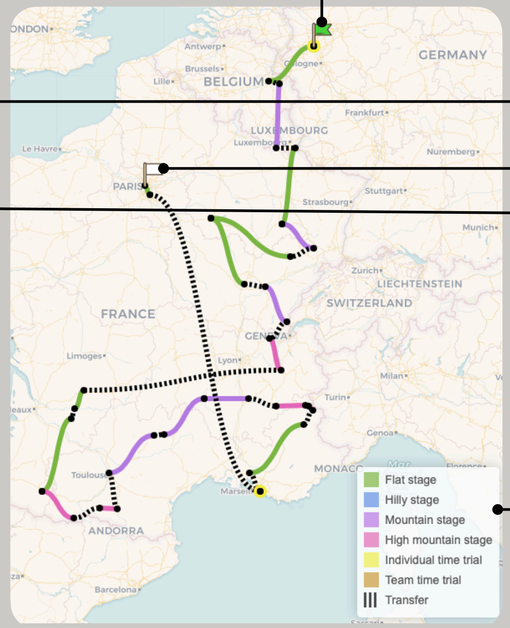
Dropdown menus that allow user to choose the edition of interest

Starting date and end date

The edition status will update itself when choose a certain edition

Statistics

Map of stages



The start of an edition

The drop down menu allows users to select how many entries that want to be displayed in the current table

The end of an edition (Note: This should be a checkered flag due to the browser used it's white now.

The search bar allows to search for riders names and/or for teams

Stages are classified by their types and different types are visualised with different colours.

This table provides detailed information about the ranking of riders, their team, time to finish an stage and time gap from the winner

General Classification:

Show 10 entries Search:

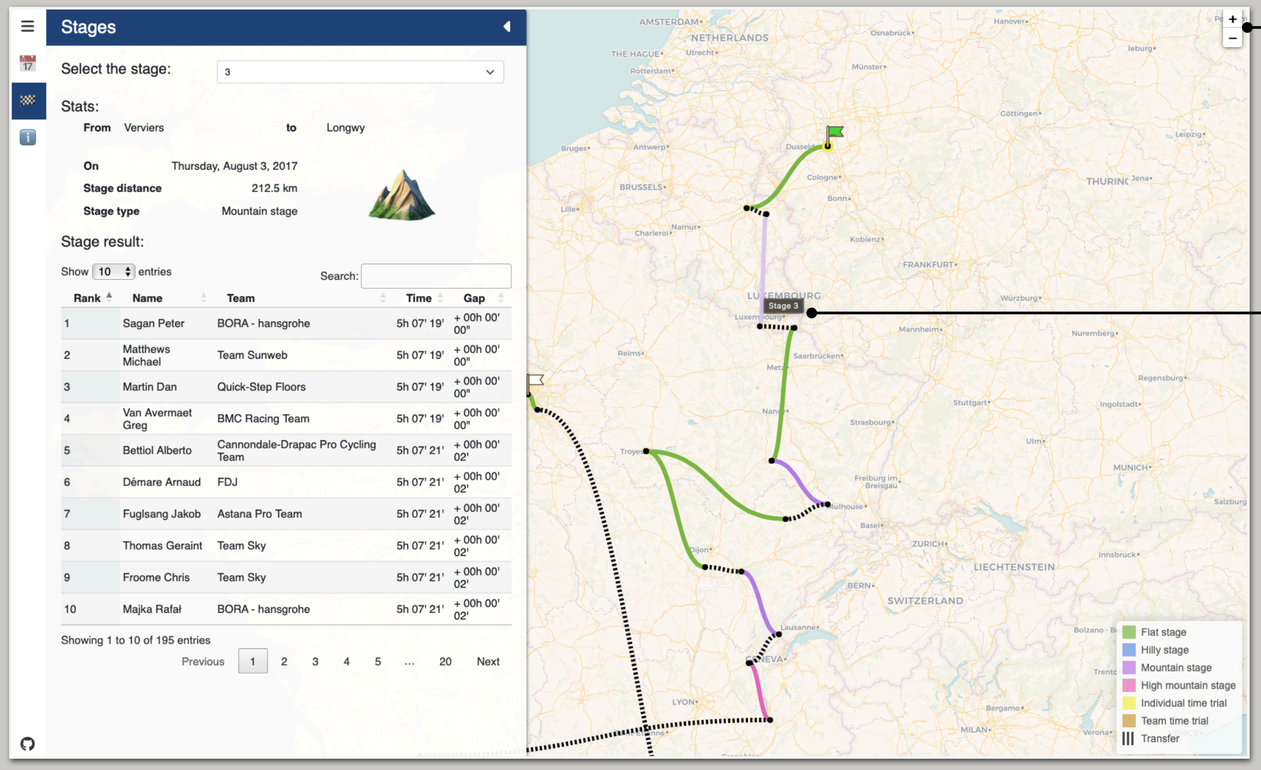
Rank	Name	Team	Time	Gap
1	Chris Froome	Team Sky	89h 04' 48'	+ 00h 00' 00"
2	Romain Bardet	AG2R La Mondiale	89h 08' 53'	+ 00h 04' 05'
3	Nairo Quintana	Movistar Team	89h 09' 09'	+ 00h 04' 21'
4	Adam Yates	ORICA-BikeExchange	89h 09' 30'	+ 00h 04' 42'
5	Richie Porte	BMC Racing Team	89h 10' 05'	+ 00h 05' 17'
6	Alejandro Valverde	Movistar Team	89h 11' 04'	+ 00h 06' 16'
7	Joaquin Rodriguez Oliver	Team Katusha	89h 11' 46'	+ 00h 06' 58'
8	Louis Meintjes	Lampre - Merida	89h 11' 46'	+ 00h 06' 58'
9	Daniel Martin	Etixx - Quick Step	89h 11' 52'	+ 00h 07' 04'
10	Roman Kreuziger	Tinkoff	89h 11' 59'	+ 00h 07' 11'

Showing 1 to 10 of 174 entries

Previous 1 2 3 4 5 … 18 Next

Implementation

As already mentioned, we do not have access to the precise route of stages. To display it we decided to draw lines between the start and end points. The technical challenges to draw dynamic looking lines and overlay them on the Leaflet map were quite big but we are quite happy with the result. Overall the scale comes across well. The final challenge in that regard were stages that started and finished at the same place. To still be able to show them we add circles around those places.
There are many towns on the planet and there are many locations in our data. We programmatically mapped those locations to coordinates in order to display them on the map. This posed issues for name conflicts and out of date names. For example, West Berlin, the starting city of the 1987 edition, was interpreted as the Canadian town instead of the German capital. To resolve those issues we had to manually go through the data and correct it as best as possible. Having our map visualization done allowed us to detect these issues quickly but this still took some time.
Since we essentially used two different but complementary datasets we had to make them coherent in time format, naming conventions and so on. Finally, some missing data, especially the winners of the various classifications were missing. We acquired this by scraping Wikipedia.



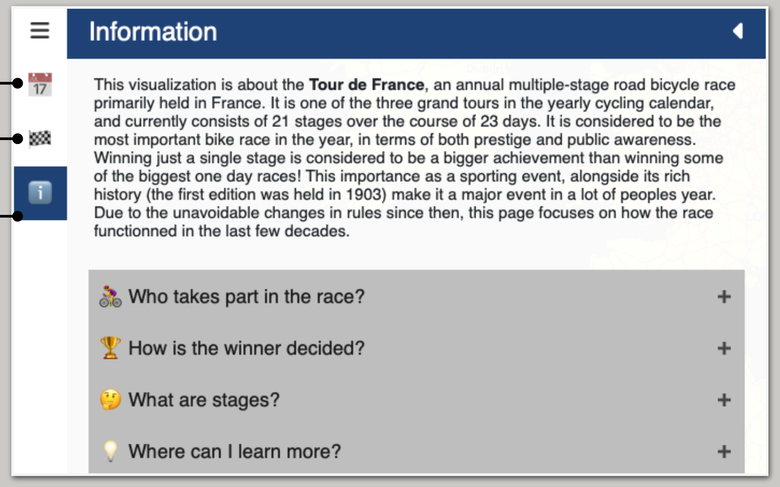
+ — button to zoom on map

Mouse over a certain stage, stage colour gets lighter and the stage number will appear on the map. On click, The stage information will update instantly.

Edition tab with a calendar icon because edition are classified by year

Stages tab with a checker flag because it presents the end of a stage

Information tab with an information icon which is standardised and crystal clear.



Our contribution

Surprisingly, despite the Tour de France being such a popular event, we did not find any visualization trying to represent it as a whole. The various documents we did find are either going into specific details and made for engaged fans like this article by Cyclingtips, dedicated to one edition like the official route overviews, or not designed to provide an interactive experience like these fan websites.

Although the maps, results and statistics of every edition of the Tour de France are readily available, for example on procyclingstats, we believe that our project adds something new. We think that our visualization does a good job of providing an overview of every edition and stage whilst also allowing us to easily see the evolution of both by using the website. All while being more attractive than a simple wall of names and numbers.

Future work

Probably the aspect of the visualization that would benefit the most of an improvement is the map. Whilst overall the spatial aspect of an edition is represented quite well, the exact distance of a stage is not. Stages that took a relatively straight line between the start and the finish will be somewhat accurate but especially the ones that start and finish in the same place will not. Having access to the exact route taken on each stage would be ideal. This could for example be achieved by scraping the Strava profiles of riders, but this approach would rule out any editions earlier than 2010. Other approaches could be implemented, like wider lines/circles depending on the distance. But then keeping the visualization readable is difficult.

Whilst quite extensive, the data used by the visualization could be improved. The exact path of a stage as mentioned above could be added, but also other interesting data like rider nationality, stage elevation gain, stage combativity prize winner, points and km classification standings after each stage, etc. Coherence of this data would require quite some work because of the difference in rules and format between the years. For example, teams might change names because of sponsors but they are essentially still the same team. More generally, modifying the visualization of each edition to best reflect the race format at that time would require a lot of work but could be worth it.

Finally, there are still many aspects of the data that are left unexplored. Speed, evolution of the classification after each race, showing data per rider, per team or per geographical location. Graphs, heatmaps or direct comparisons could all be added. All of this can be used to enable a more clear comparison between stages or editions.

Peer Assessment

Maxime Jan: Design of overall visualization concept. Exploratory data analysis. Set up map and sidebar skeleton. Display stages on the map. Add result tables to the sidebar. Screencast.

Jonas Konrad: Design of overall visualization concept. Data cleanup and processing. Display stages on the map. Interaction/coherence of stage selection between sidebar and map. Information sidebar. Implementation of website design. Website polishing. Process book text content.

Nana Tian: Implementing the status information in the stages and edition tab. Cover the visual design of the process book, part of text and merge the text together.