CPSC 583 - W2021

Jordan Lee (30002218)

P2 - Sketching

Sketchable Data Subsets

Subset 1: Stage Number, Distance, and Date

The focus for subset 1 was to show how stage distances of the Tour de France changed over time. This subset contains over 2000 rows and 3 columns, and every year of the Tour in this dataset has data in this subset. Because not every year of the Tour has the same number of stages (especially the earlier years), there is "missing data". The varying number of stages, as well as the changing distance covered by each stage, can provide insight to the difficulty of this sporting event and how it has changed over the years.

However, this subset is still too large to sketch in its entirety, so sketches were based on a smaller subset. Certain sketches showed data for a few years at a time, or a limited selection of stages. I also explored the idea of visualizing a single year, and comparing it to the rest of the data through averages, minimum/maximum values, etc.

See ./subset1.csv

Subset 2: Origin/Destination City, and Date

Subset 2 includes the origin and destination cities, and dates. This subset contains data for every single tour (recorded in this dataset). Although date is included, there is less emphasis on how they change over time and more on the year that they were involved. Being a host city can bring prestige and an economic boom to the area, and visualizations of this subset can be helpful for further analysis.

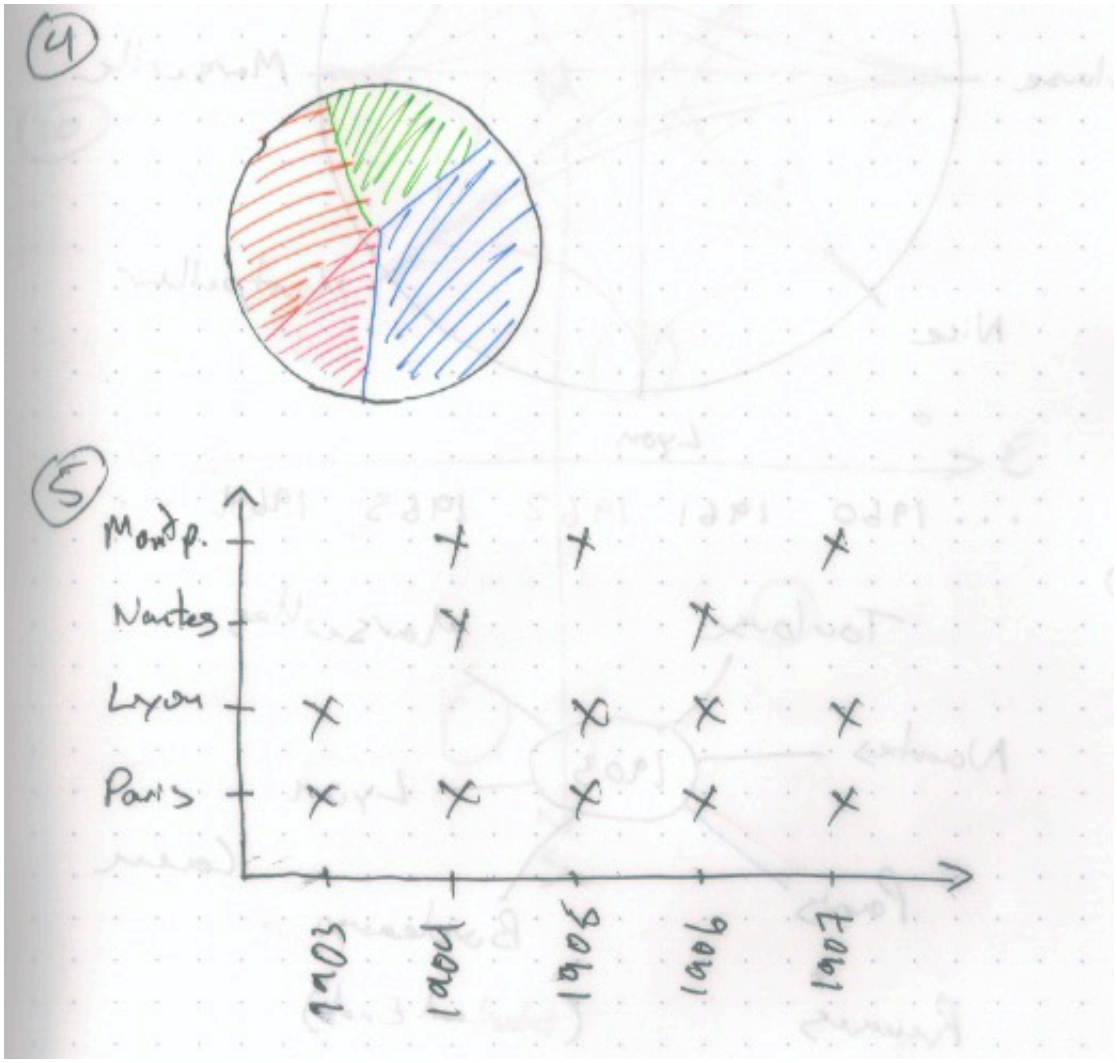
This subset is of similar size to Subset 1, but due to the categorical nature of the data I found it easier to include the entire subset in my sketches. Some sketches visualize only one year in comparison to the data, and some sketches do not include date at all and focuses on the frequency of popular cities included in the Tour.

See ./subset2.csv

Design Direction

The refined sketches reflects the more common and intuitive ways to represent the data. For example, distances over time are commonly expressed as line or bar charts, and cities are intuitively visualized on a map. There are novel ideas introduced in the initial sketches, but the traditional visualizations proved to be more effective and expressive, and were more likely to be chosen to be developed further.

Not all sketches (both initial and refined) visualize the same information for the same dataset. For instance, one sketch for Subset 2 represents the overall number of times that a city hosted a stage of the Tour de France, but another sketch for the same subset shows the specific year that the city was an origin or destination city. As expected, the frequency visualizations are generally more compact and more easily represented the entire subset of data compared to the specific year-to-year visualization, which only include a few years in the sketch. Even though some of the visualizations with a limited scope are effective, when they are scaled to include the entire subset it can be overwhelming to comprehend because of the large amount of small multiples.

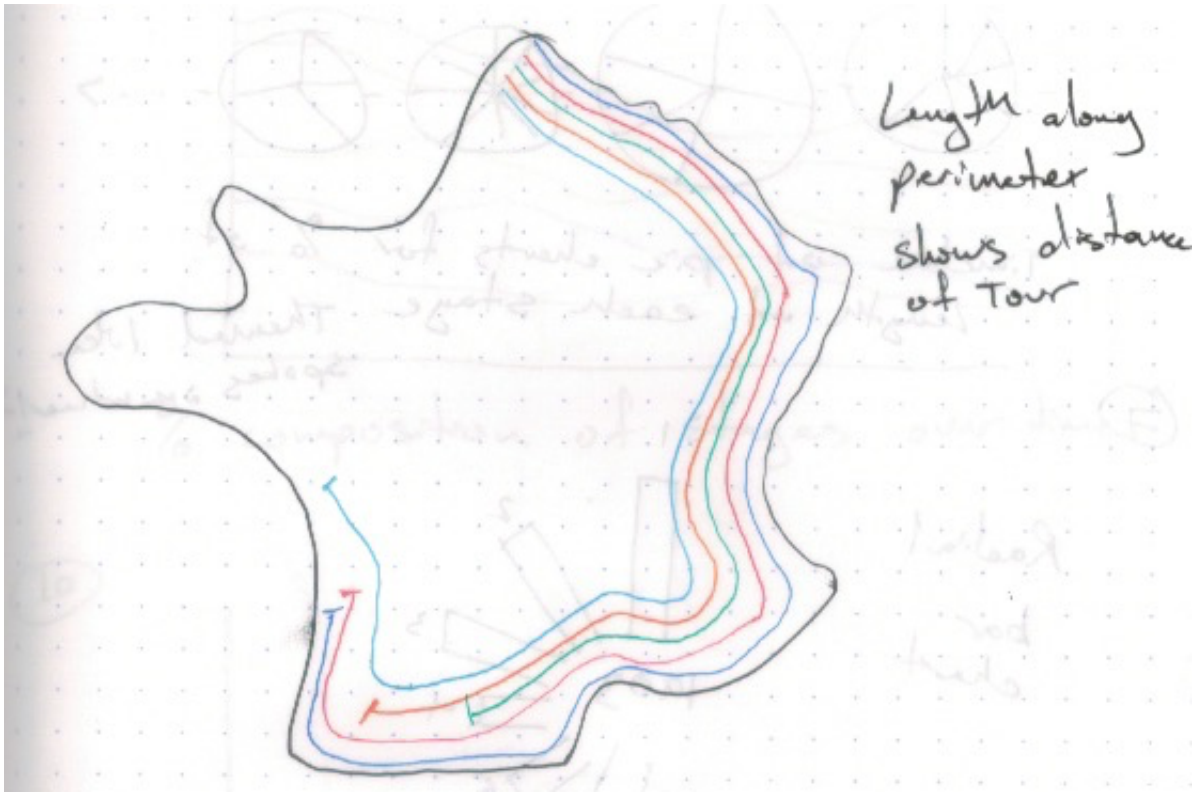


Process and Results

The two subsets were created with a clear goal in mind, which drove the results of the sketches. I approached the visualizations first with immediate ideas on how the information could be represented intuitively. Next, I explored ideas that were commonly used to show similar data. Finally, if I did not have ten initial sketches yet, I would try visualizations that were not traditionally used if they could represent the data.

Since some of the sketches are based on a subset of the subset, ideas that could not scale to represent the entire dataset were not explored further. To refine the sketches, I would modify ideas from the initial sketches by changing the type of encoding, or apply the same visualization to a different subset.

Experimenting with colour made it apparent that there were not enough distinguishable colours to represent the big range of variables. As an example, there are at most 21 stages but there are less than 21 variations of colour that can be easily distinguished. Rather than assigning a colour to every single value, I used colour to highlight certain information from the rest or to represent a range of values.

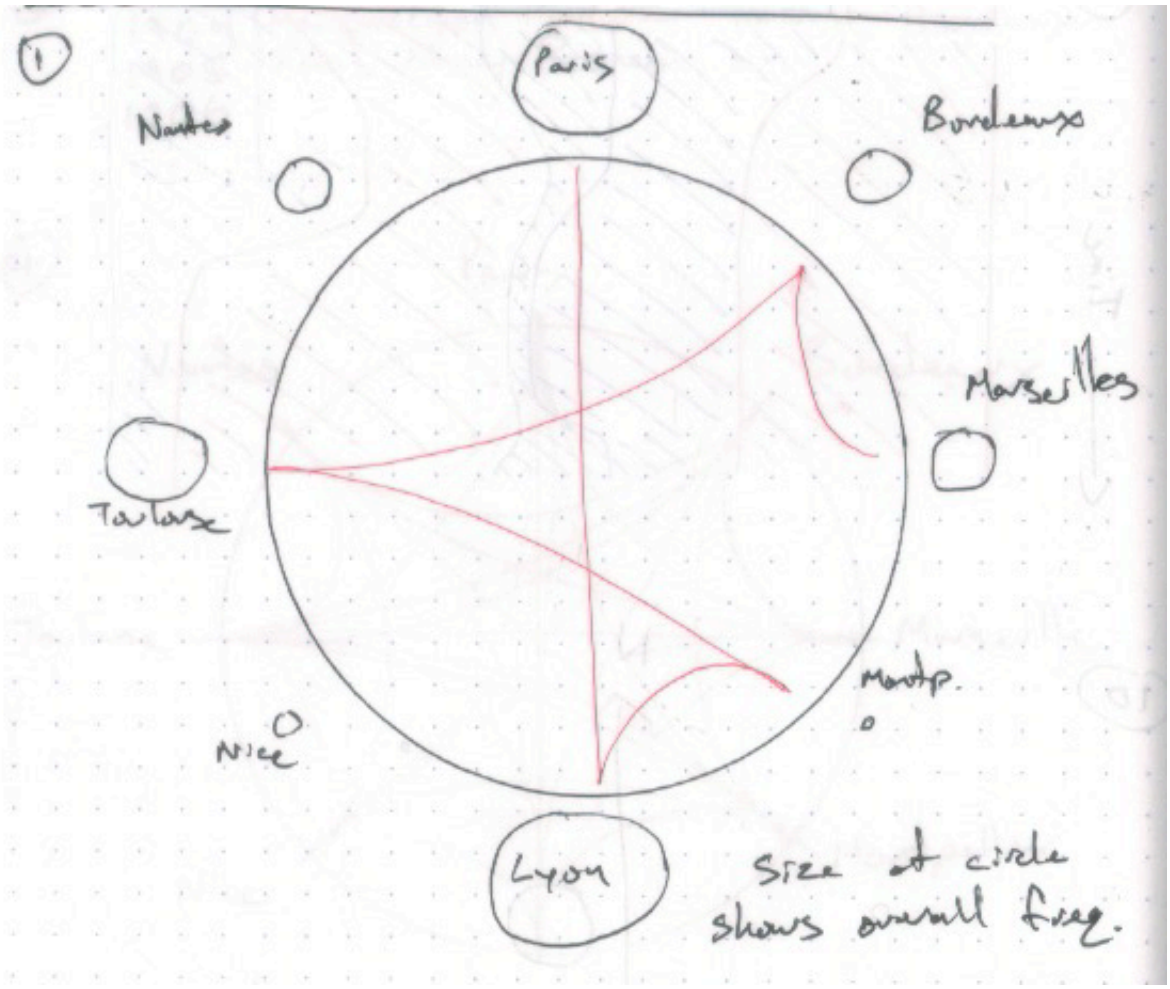


As well, the visualizations focused only on the subset and may not work as well if the subset was expanded to include other variables. The refined sketches have ideas that are specialized to represent that information, and introducing another column to the subset is difficult.

Conclusion

From the sketches done, showing stage numbers and distances over years is likely best with a line or bar chart. Representing the large amount of data all at once in an easily comparable and comprehensible manner is a challenging task, and this traditional visualization is a compact way of accomplishing it. For Subset 2, cities are commonly shown on a map, and by leveraging this familiarity an effective visualization can be created.

On the other hand, non-traditional visualizations are not without their strengths. I found that the circle representation for refined Subset 2 sketches was particularly effective, in that it was able to show both overall data as well as a specific year's information at the same time. Furthermore, it can be easily adapted to include even more information, such as the order in which cities are visited of a given year's Tour de France.

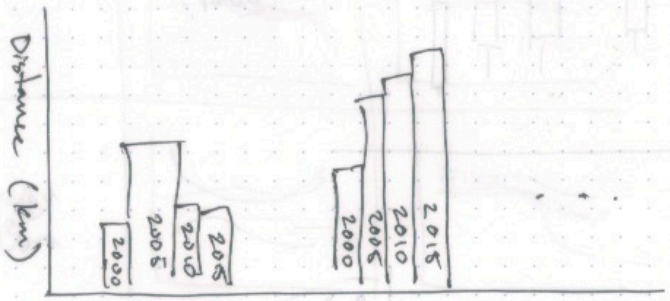


Appendix

P2 - Sketching

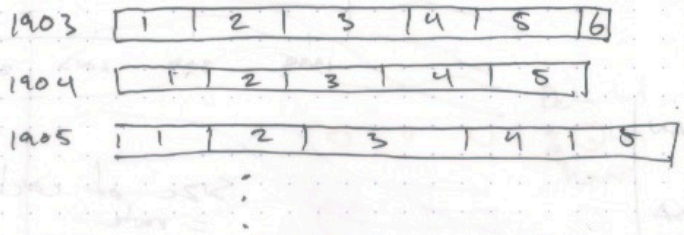
Subset 1 - Stage Number and Distance, and Date

(1)



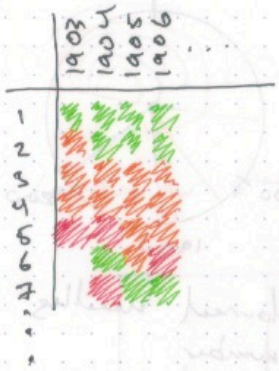
Bar chart comparing stage distances

(2)



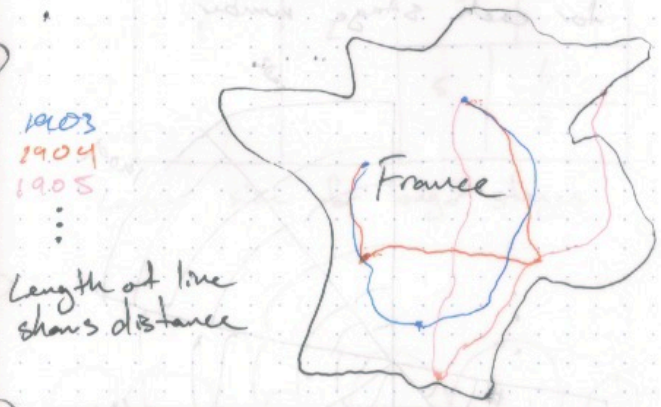
Size of block represents stage length

(3)



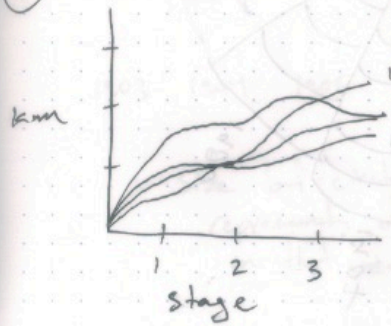
Colour Ranges for distances

(4)



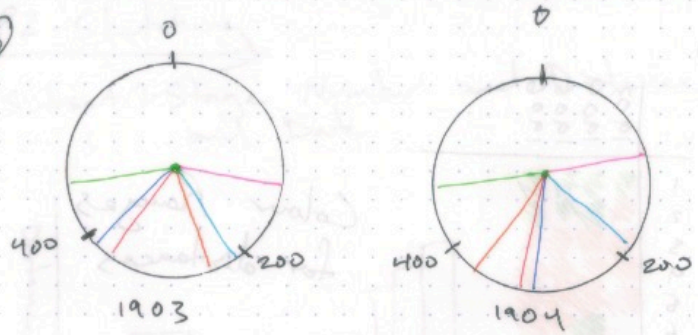
Length of line shows distance

(5)



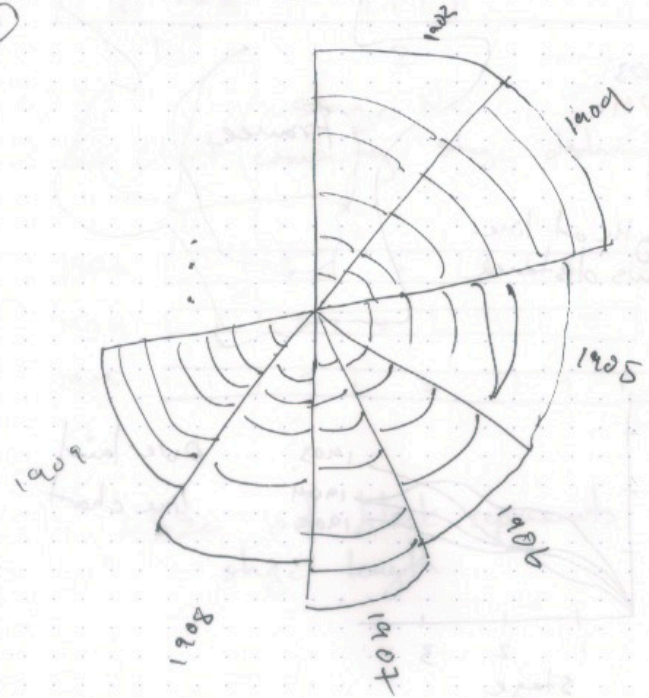
Overlaid line chart

(6)



"Gauges", different coloured needles for each stage number

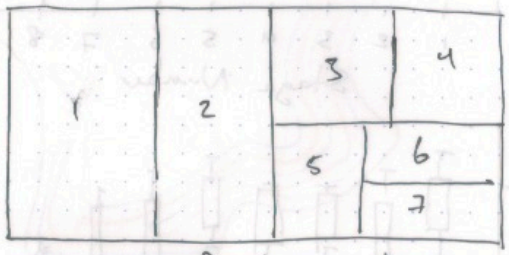
(7)



(8)

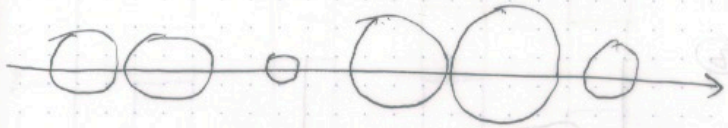
Colours for stages, length for distance.

(9)



Larger size for longer stages

(10)

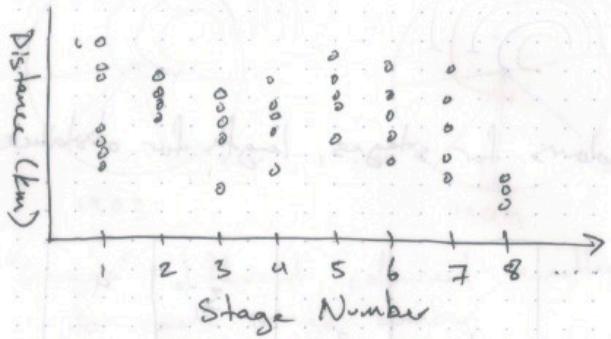


1903 1904 1905 1906 1907 1908 ...

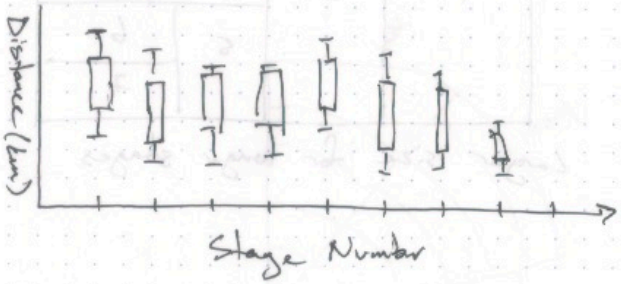
Size of circle on timeline represents distance covered in that year's tour.

Subset 1 - Refined Sketches

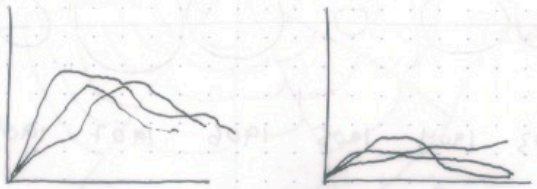
(1)



(2)



(3)



Pre - 19xx

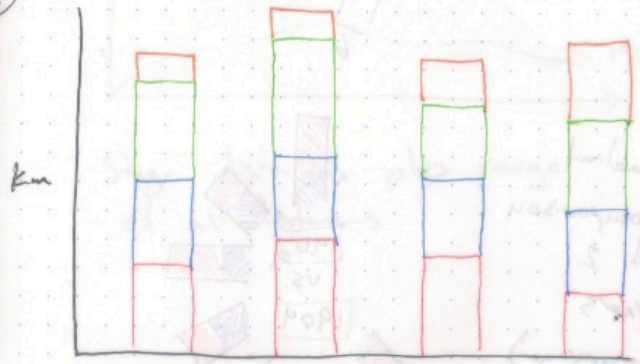
Post - 19xx

"Turning Point"

(4)

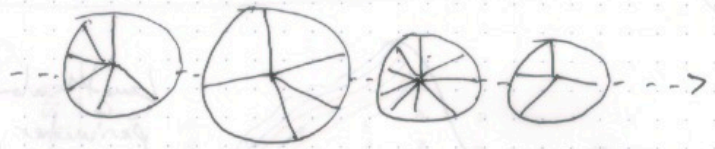
Length along perimeter shows distance of Tour

(5)



Stacked Bar Chart with colours denoting stages

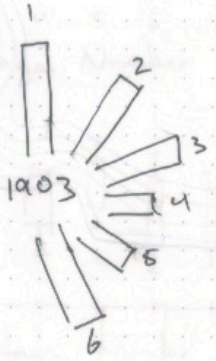
(6)



Timeline of pie charts for % of length of each stage. Themed like spokes on wheel?

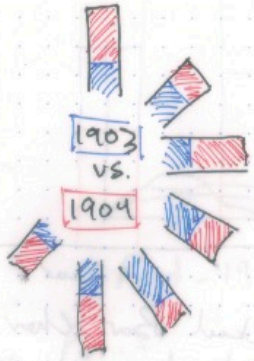
(7)

Radial bar chart

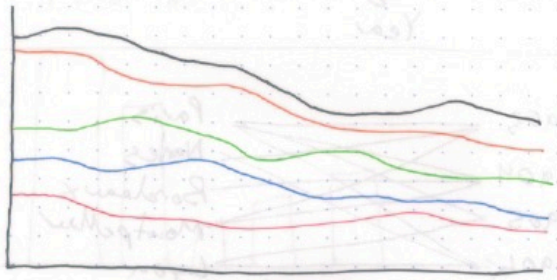


(8)

Radial comparison of 2 years (or average)

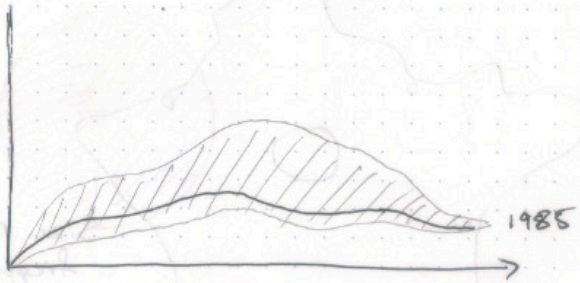


(9)



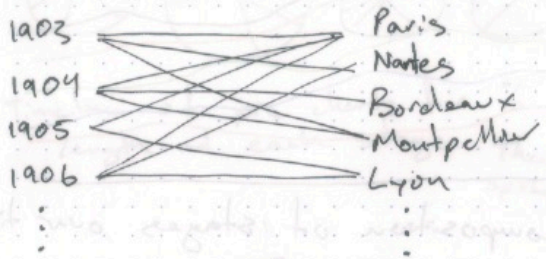
% composition of stages over time

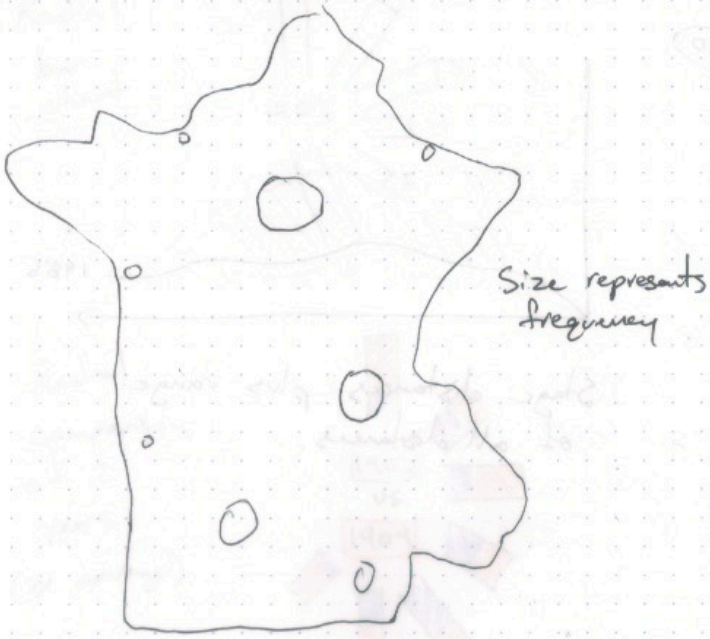
(10)

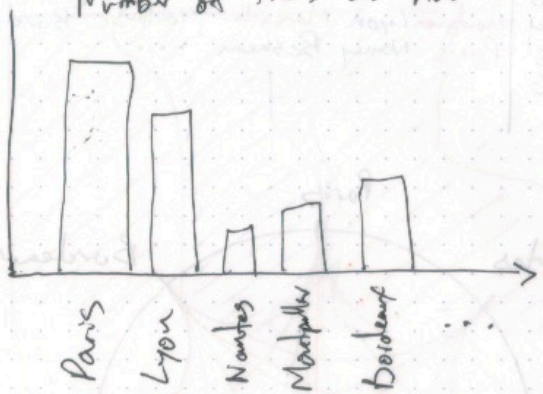


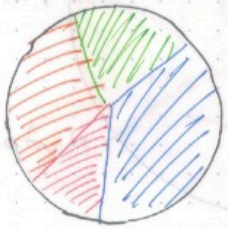
Stage distances plus range of all distances.

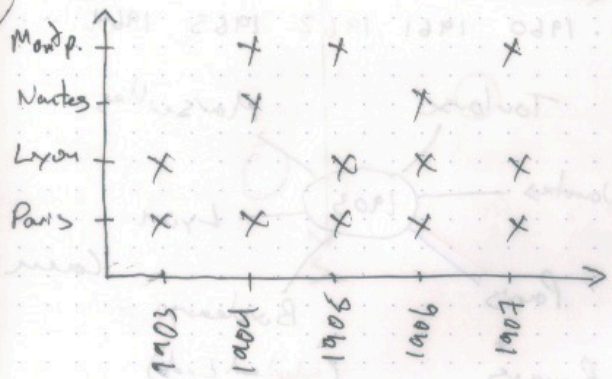
Subset 2 – Origin/destination cities, and Year

① 

② 

③ 

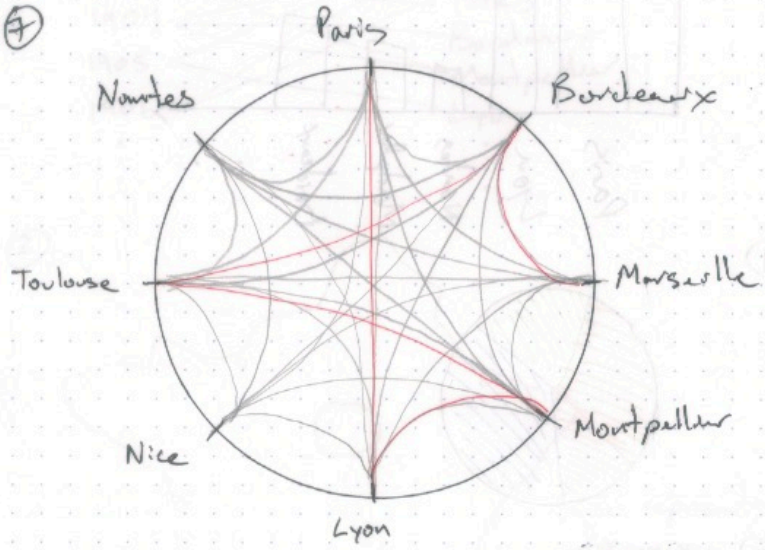
④ 

⑤ 

(6)

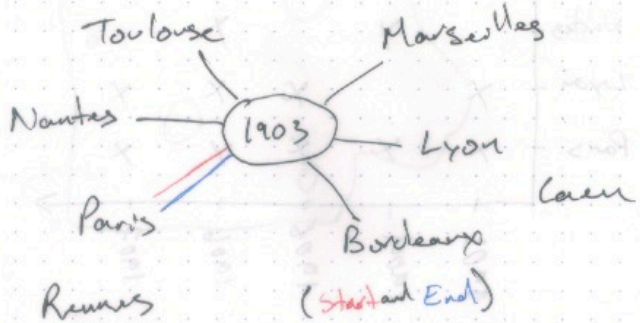
1903	Paris	Lyon	Marseille	Toulouse	Bordeaux
1904	Montpellier	Lyon	Marseille	Toulouse	Bordeaux
1905	Paris	Nancy	Besacon	...	
1906					
⋮					

(7)

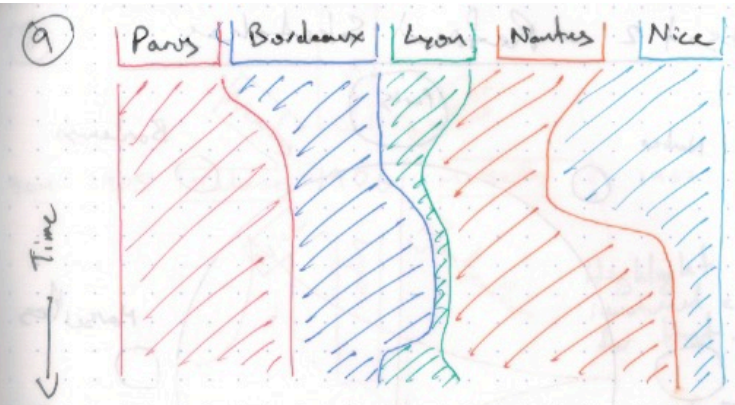


... 1960 1961 1962 1963 1964 ...

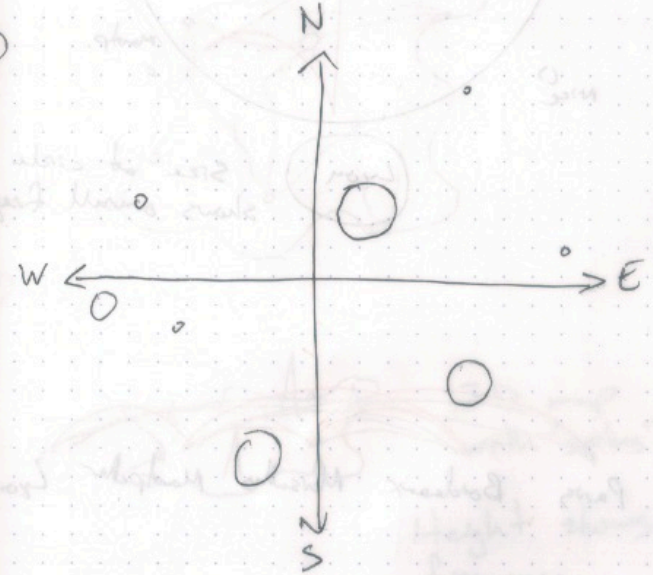
(8)



(9)

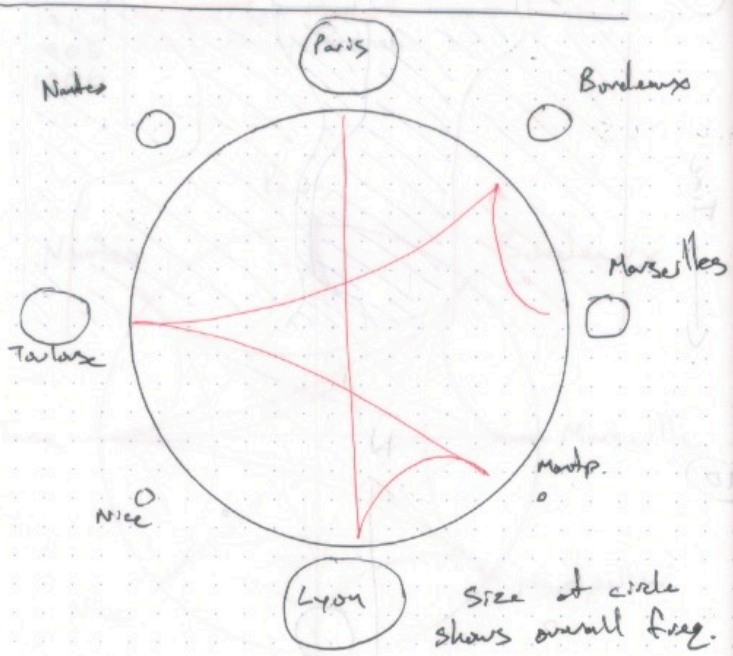


(10)



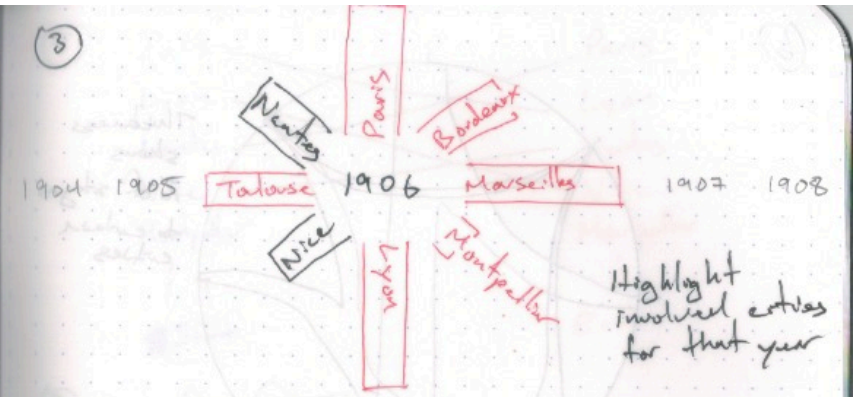
Subset 2 - Refined Sketches

1

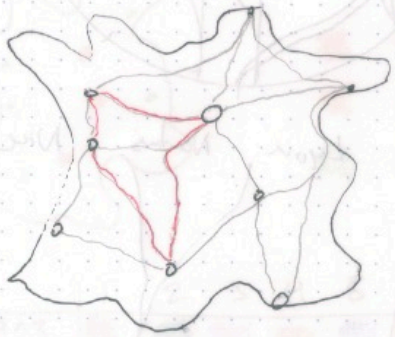


2

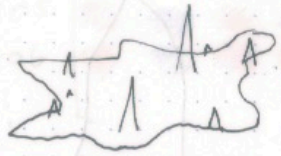
3



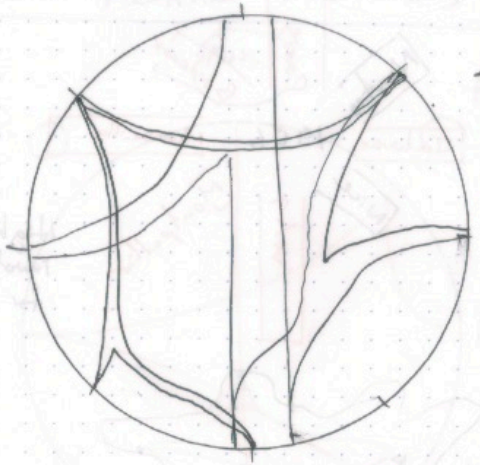
4



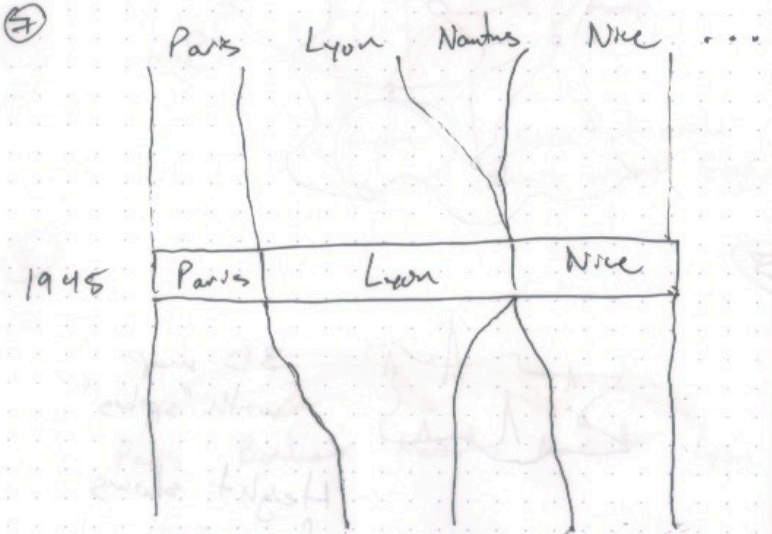
5



(6)

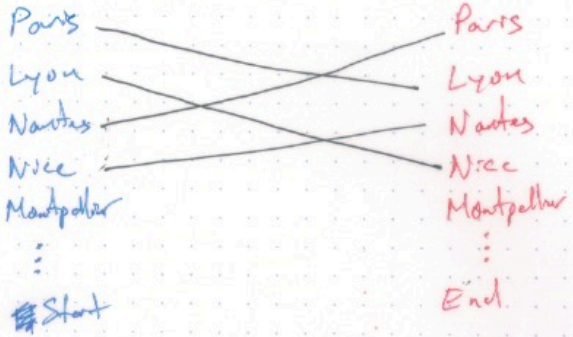


(7)



Year "slider" to show details of that Tour

(8)



(9)

(10)

Colours to represent cities

BUDDY Group Contract

Date: February 20th, 2021

Expectations from Team Members

Expectation	Consequence if expectation not met
We'll meet as a team every week at: **Friday 12:00pm for 1 hrs**	Team meeting will be Friday at 12, after our tutorial time. If this does not work then team members need to notify the other and reschedule at a time that works for both group members.
Be on time and prepared for team meetings. (Both in and out of class time)	Team members are expected to meet on time, and communicate with the team/reschedule for another time if the regularly scheduled meeting does not work.
Follow through on commitments made to the team. (These are likely to be minimal as your projects are individual, but may include committing to giving thoughtful feedback on members' work.)	Make up missed commitments for the next meeting. Keep continuous contact with other team members if you expect to miss any commitment so there are no surprises for other members.
Contribute to the team voluntarily. (These contributions may be ideas, questions, code, organizing meetings, managing code repository, creating charts for the report, etc.)	Team members should always put their best foot forward in trying to help the other members throughout the term. Team members are there to help each other succeed so they should play to each other's strengths and always work for both members' success. If team members are not putting any effort into any ideas except their own then they should be called out and they are expected to help moving forward.
Welcome and invite contributions by other team members.	Any team members not allowing contributions from others should be reported to the professor for causing a toxic work environment and if they do not improve then they should be removed from the group.
How you define the line between support and plagiarism. (helpful to play out a scenario, e.g. a buddy group member is stuck with a block of code)	Team members can provide ideas to other members but it turns into plagiarism when no input or changes are made from the receiving side and instead they just take the idea and roll with it. Ideas should spark content but there should always be more contribution from the owner of the project.
Each team member will talk for at least 10 minutes, and at most 30 minutes during a 1-hour meeting.	Team members will continuously check in with the other members to see if they have input and always make sure that they are not either dominating the conversation, or not talking enough.

If you've read and agreed with this contract, add you name here:

Jordan Lee

Jason Lyster