F1 2011 Race Results

Project Description

Kendrick Cancio (kdc57)
Saloni Joshi (lsj38)
Angeline Pinilla (ang58)

Overview

Formula 1 is a sport rich in data representing the excitement and uncertainty of a race. Unfortunately, much of this data is reduced to just the finishing times for each driver, completely ignoring the events that transpired throughout. In this visual, we aim to capture changes in overall driver position from lap to lap while also showing events where drivers are forced to retire from a race.

DATA

Collection -
We aggregated data from two sources:
http://www.f1datajunkie.com/p/data.html
http://ergast.com/mrd/

The first contains csv files of the 2011 Formula 1 races where for each driver and lap, their time gap to the race leader is calculated. Because we aimed to create individual driver path elements, we needed to convert the csv into a json file that could represent the drivers as objects with lap objects nested within them for each race. To do this, the following python code was written:

```
import numpy as np
import pandas as pd
import json

Race = pd.read_csv("F1_2011_Spain_TimingData-Comprehensive.csv",sep=",")
Race = Race[[0,2,8,13,12]]

def pdtodict(df):
    Race_Dict = {}
    drivers = set(Race["driver"].tolist())
    for d in drivers:
        Race_Dict[d] = {}
        temp = Race[Race["driver"] == d]
        laps = set(temp["lap"].tolist())
        for l in laps:
```

```
        lap_str = str(l)
        temp2 = temp[temp["lap"] == l]
        Race_Dict[d][lap_str] = temp2.values.tolist()[0][1:]

    return Race_Dict

x = pdtodict(Race)

with open('Spain.json', 'w') as outfile:
    json.dump(x, outfile)
```

The second website contained additional information regarding the finishing status of each racer along with their nationality. We were able to request json files directly from the site to use in our visualization.

Cleaning the data proved rather arduous at times. Across our two sources, there was no consistency in terms of naming conventions for the drivers. We were forced to parse the driver names to keep them consistent across sources. We also encountered 1 instance in which the driver name “Kobayashi” was misspelled as “Kobyashi”.

VISUAL ELEMENTS & SCALES

Driver Paths -
X scale - Lap
Y scale - Driver Position
Our visual element is a graph representing all the laps in the selected 2011 F1 race. When a race is selected, paths begin to be drawn from the left side of the screen signifying the start of the race. For each lap, a driver’s overall position is plotted; these points form the path that represents the changes in position throughout the race.The speed with which each line is drawn is linearly dependant on the finishing order of the racers so that the first place finisher reaches the right hand side of the screen before the second place finisher and so on. Although the finishing speed isn’t scaled to the actual relative finishing times of the race among the drivers, it serves to train the user to expect the fastest finishers to be listed from top to bottom on the right hand side of the screen.

An additional interactive element to the paths is the ability to hover over them to dim the other paths and highlight the selected driver’s name. This helps the user navigate parts of the graph where lines intersect heavily, allowing them to focus on just one driver.

Driver Names -
Y scale - Driver Finishing Position
On the right hand side of the visualization are listed the driver names in their finishing order along with a small flag representing their country of origin and their finishing status. A red circle indicates the driver did not finish the race due to a mechanical failure or collision. An orange circle indicates the driver completed the race at least a lap behind the 1st place finisher. A green circle indicates the driver finished on the same lap as the leader.