Rocket Powered Racecars

STAT 385 FA2018 - Team Rocket Power

Emma Plansky (plansky2)
Siddharth Ahuja (sahuja7)
Alex Wang (apwang2)
Waseem Ebrahim (webrah2)

December 18, 2018

Abstract

Formula One's popularity has only grown since it formally became a sport back in 1950.
Between 1950 and today, Formula One has evolved significantly, making it difficult to compare car performance, driver performance,safety and other factors of the sport over the years.In our project, we will be using data from 68 seasons of Formula One (1950 - 2017) to better understand the evolution of the sport.

Contents

1 Introduction

Evidence for Rule Changes and Evolution of Formula One:

The regulating body of Formula One, the Federation Internationale de L'Automobile (FIA), changes several regulations for the sport on a year-to-year basis. This has resulted in several changes to how Formula One cars perform. The specific differences in regulation are published by the FIA every year. We can also see the differences in regulation by simply looking at how a Formula One car from 1950 compares to one from today.

These regulations have also changed how the points for a race are allocated. This can be seen by looking at Point Scoring Systems table in the appendix [Figure 1 in Appendix]. If the current points system was in place throughout the years it may have resulted in different drivers and teams winning championships. Different years also had a different number of teams participating in the sport which changed how hard it is to score points (if there are fewer teams participating it is easier to achieve a points scoring position).

Addressing the Problem:

As stated above, the problem created by all of these changes in the sport is that it is hard to compare the drivers and teams in the sport over the years from the ones today. To address this problem we normalized the results of races over the 68 years as accurately as possible. The results were normalized by taking into consideration changes in the sport. For example, we used the race results data and apply the current points system to it. Then we determined which drivers and teams would have won the championships with the current points system applied.

Formula One fans (who are the target audience of this project) would interested in these results because they will be able to better compare teams and drivers who have participated in the sport.

Data:

The data for this project is from the 'Ergast Database'. The database contains 13 datasets with information on different aspects of Formula One (e.g. race results, lap times, qualifying results, driver information). Most of these datasets have several hundreds of observations. These datasets were merged and analyzed as needed to normalize the data over 1950 to 2017. Refer to the 'Methods' section below to see how the data was processed to achieve the results we need. The variables present in each dataset have been specified in Figure 2 of the Appendix. A small subset of all of the datasets have been shown in Figure 3 of the Appendix.

Relation to STAT 385:

This idea matches with this course's focus on statistical programming because we are manipulating several datasets and extracting information from them.

2 Related Work

As stated previously in our introduction, Formula One racing has grown immensely in the last 68 seasons, making it an extremely interesting topic to analyze. After some further research into our topic, we found a few interesting sources that are related to what we were trying to accomplish. Although we were able to find sources to help us gain further knowledge into our topic, we were not able to find any research that had the same goal as our project. As Formula One racing grows in popularity, more and more research is being conducted on the efficiency of cars and creating models to decrease lap times. Our project, however, largely focuses on normalizing race results over the past 68 seasons in order compare results and see how Formula One races have evolved over time on a deeper level.

We first looked at Simulating Formula One Race Strategies, conducted by Claudia Sulsters from the Vrije University in Amsterdam. The goal of her research was to create a simulation model that could be used by Formula One drivers to enhance their performance. By using her model, racers are able to simulate different race strategies to determine how these strategies will affect their result. In Sulster's research, she used data from four races in the 2016 season. While Sulsters work is extremely interesting and useful to racers, it does

not take into account the ever changing rules and regulations of the sport. She uses only data from 2016 and since the engineering of the cars, point systems and other rules will change, her model will eventually become out of date in the coming years. For our project, we will standardize the results in order to compare every year starting from 1950 when the sport became official.

In addition to Sulter's research, we also found Formula 1 Race Car Performance Improvement by Optimization of the Aerodynamic Relationship Between the Front and Rear Wings by Unmukt Rajeev Bhatnagar, a graduate student in aerospace engineering at Pennsylvania State University. The research done by Bhatnagar is relevant to our project since he looks at how the design of cars has changed within the 27 seasons. His main goal is to create a vehicle that is more aerodynamic to decrease lap times. In relation to this research conducted by Bhatnagar, our project is vastly different as we do not aim to create a more efficient car, but to analyze previous efficiencies by normalizing results for better comparison.

The last source that we found, InfoZoom - Analysing Formula One racing results with an interactive data mining and visualisation tool, is a report on a visualization tool used for Formula 1 racing. This interactive data analysis tool is called InfoZoom. The paper was written by Michael Spenke and Christian Beilken from the German National Research Center for Information Technology. The tool was created to explore the statistics of each individual driver. The information that is available from this tool only includes the years 1978 to 1998. The tool displays pictures of the drivers, information on what type of car the driver was most successful with, and other race statistics for each specific driver. The goal of this visualization is to make the user more comfortable with the data, however, it does not dive deep into analyzing the data that is displayed.

The idea that our team has come up with is very useful for Formula One enthusiasts as we aimed to make comparisons for the 68 seasons available to those who are interested. Due to the large amounts of data that has been collected for Formula One, there is a lot of research that has been conducted. However, our project brings in another element to the analysis which has not been looked at before.

3 Methods

The goals of the project discussed in the 'Introduction' were achieved by completing the following steps:

Reading the data into R:

To begin our project, we first had to read the data, which were all CSV files, into R. We started with a total of 14 datasets which were uploaded using the function read.csv(). After uploading the data, we then moved on to data cleaning.

Data Cleaning:

The first step we took in cleaning the data was adding variable names to the columns and all 14 tables. After those were added, we started looking at the levels within each of the variables. As our data was pretty clean to begin with, the only further cleaning that needed to be done was merging certain datasets as required for each of the visuals. This was completed throughout the project as the visuals were created.

Data Visualizations:

In order to create the table visualizations seen in our Shiny application, we used dplyr join functions to merge together the different tables. In this process we also used the 'subset' function to drop any unnecessary variables that were not needed in our final tables.

In our Shiny application, we also wanted to include barplots and other graphs in order to better display what can be seen in the tables. To create these visuals, we used ggplot2, a package created by Hadley Wickham.

Presenting the data:

As mentioned above we created a Shiny application, which was obviously created using the shiny package.

4 Results

The important tables and visuals from our project can be found in the shiny application itself. Below are some examples of these important tables (only the first few observations are shown):

Year	Race Name	Driver First Name	Driver Last Name	Constructor Name	Finishing Position	Original Points	New Points
2008	Australian Grand Prix	Lewis	Hamilton	McLaren	1	10.00	25.00
2008	Australian Grand Prix	Nick	Heidfeld	BMW Sauber	2	8.00	18.00
2008	Australian Grand Prix	Nico	Rosberg	Williams	3	6.00	15.00
2008	Australian Grand Prix	Fernando	Alonso	Renault	4	5.00	12.00
2008	Australian Grand Prix	Heikki	Kovalainen	McLaren	5	4.00	10.00
2008	Australian Grand Prix	Kazuki	Nakajima	Williams	6	3.00	8.00

New Points Table

Driver First Name	Driver Last Name	Average Points per Race
Carlo	Abate	0.00
George	Abecassis	0.00
Kenny	Acheson	0.00
Andrea	de Adamich	1.26
Philippe	Adams	0.00
Walt	Ader	0.00
Kurt	Adolff	0.00
Fred	Agabashian	2.50
Kurt	Ahrens	0.00
Alan	Brown	2.22
Christijan	Albers	0.24
Michele	Alboreto	3.61

Figure 1: Average Points per Race Table

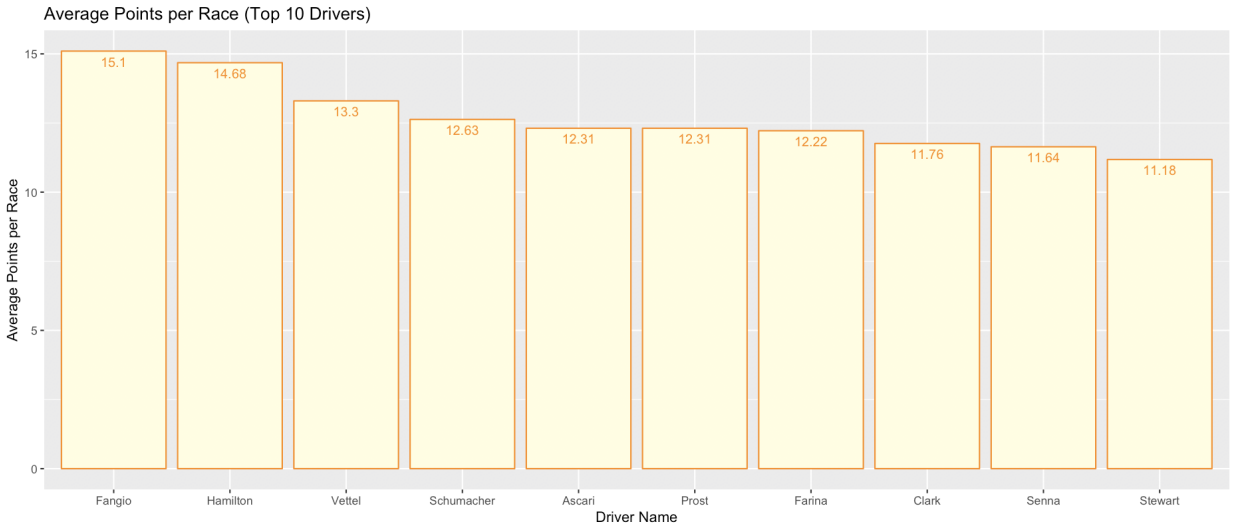

Figure 2: Average Points per Race Visual

5 Discussion

The overall purpose of our project was to compare drivers and teams in Formala One across the different eras of the sport. In order to achieve our outcome, a variety of variables were analyzed to determine whether or not they would be appropriate for the comparison of drivers and teams.

As mentioned above, one of the ways that we compared drivers/teams in Formula One is by normalizing the number of points that were earned to the modern point system. The purpose of this is so that Formula One enthusiasts can compare their favorite drivers from previous eras of the sport to current divers.

From this analysis and the visuals we created we were able to determine that teams like Ferrari, McLaren, Williams and Mercedes have been dominant in Formula One. Drivers like Schumacher, Vettel, Alonso and Hamilton also appear regularly as some of the top drivers to have raced in the sport.

6 Conclusion

Formula One has evolved and grown tremendously as a sport over the last several decades. As a result of its growth, over the total 68 seasons, the rules and regulations have changed with almost every passing year. We strived to make interactive visuals that show the evolution of the sport and compare the results of these races in a normalized manner. For those who enjoy watching Formula One and seeing how the cars and drivers have changed, this project gives them an opportunity to accurately make comparisons over the years.

Our Shiny Application provides enthusiasts with a chance to compare their favorite drivers and teams through an interactive model. For further information on how our application can be utilized, our demo video will provide a clear explanation.

7 Appendix

Figure 1

Seasons	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Fastest Lap	Best Results Counted Towards Drivers' Championship	Best Results Counted Towards Constructors' Championship
1950-1953	8	6	4	3	2						1	4	NA
1954	8	6	4	3	2						1	5	NA
1955	8	6	4	3	2						1	5	NA
1956-1957	8	6	4	3	2						1	5	NA
1958	8	6	4	3	2						1	6	6
1959	8	6	4	3	2						1	5	5
1960	8	6	4	3	2	1						6	6
1961	9 (D) 8 (C)	6	4	3	2	1						5	5
1962	9	6	4	3	2	1						5	5
1963-1965	9	6	4	3	2	1						6	6
1966	9	6	4	3	2	1						5	5
1967	9	6	4	3	2	1						9 (5 from first 6, 4 from last 5)	9 (5 from first 6, 4 from last 5)
1968	9	6	4	3	2	1						10 (5 from first 6, 5 from last 6)	10 (5 from first 6, 5 from last 6)
1969	9	6	4	3	2	1						9 (5 from first 6, 4 from last 5)	9 (5 from first 6, 4 from last 5)
1970	9	6	4	3	2	1						11 (6 from first 7, 5 from last 6)	11 (6 from first 7, 5 from last 6)
1971	9	6	4	3	2	1						9 (5 from first 6, 4 from last 5)	9 (5 from first 6, 4 from last 5)
1972	9	6	4	3	2	1						10 (5 from first 6, 5 from last 6)	10 (5 from first 6, 5 from last 6)
1973-1974	9	6	4	3	2	1						13 (7 from first 8, 6 from last 7)	13 (7 from first 8, 6 from last 7)
1975	9	6	4	3	2	1						12 (6 from first 7, 6 from last 7)	12 (6 from first 7, 6 from last 7)
1976	9	6	4	3	2	1						14 (7 from first 8, 6 from last 7)	14 (7 from first 8, 6 from last 7)
1977	9	6	4	3	2	1						15 (8 from first 9, 7 from last 8)	15 (8 from first 9, 7 from last 8)
1978	9	6	4	3	2	1						14 (7 from first 8, 7 from last 8)	14 (7 from first 8, 7 from last 8)
1979	9	6	4	3	2	1						8 (4 from first 7, 4 from last 8)	All
1980	9	6	4	3	2	1						10 (5 from first 7, 5 from last 7)	All
1981-1990	9	6	4	3	2	1						11	All
1991-2002	10	6	4	3	2	1						All	All

Seasons	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Fastest Lap	Best Results Counted Towards Drivers' Championship	Best Results Counted Towards Constructors' Championship
2003-2009	10	6	4	3	2	1						All	All
2010-present	25	18	15	12	10	8	6	4	2	1		All	All

Table 3: Circuit Data

circuitId	circuitRef	name	location	country	lat	lng	alt
1	albert_park	Albert Park Grand Prix Circuit	Melbourne	Australia	-37.84970	144.96800	10
2	sepang	Sepang International Circuit	Kuala Lumpur	Malaysia	2.76083	101.73800	\N
3	bahrain	Bahrain International Circuit	Sakhir	Bahrain	26.03250	50.51060	\N
4	catalunya	Circuit de Barcelona-Catalunya	Montmeló	Spain	41.57000	2.26111	\N
5	istanbul	Istanbul Park	Istanbul	Turkey	40.95170	29.40500	\N
6	monaco	Circuit de Monaco	Monte-Carlo	Monaco	43.73470	7.42056	\N
7	villeneuve	Circuit Gilles Villeneuve	Montreal	Canada	45.50000	-73.52280	\N
8	magny_cours	Circuit de Nevers Magny-Cours	Magny Cours	France	46.86420	3.16361	\N
9	silverstone	Silverstone Circuit	Silverstone	UK	52.07860	-1.01694	\N
10	hockenheimring	Hockenheimring	Hockenheim	Germany	49.32780	8.56583	\N

Table 4: Constructor Results Data

constructorResultsId	raceId	constructorId	points	status
1	18	1	14	\N
2	18	2	8	\N
3	18	3	9	\N
4	18	4	5	\N
5	18	5	2	\N
6	18	6	1	\N
7	18	7	0	\N
8	18	8	0	\N
9	18	9	0	\N
10	18	10	0	\N

Tables in Formula 1 Database
circuits
constructorResults
constructorStandings
constructors
driverStandings
drivers
lapTimes
pitStops
qualifying
races
results
seasons
status

10 observations from each table as examples:

Table 5: Constructor Standings Data

constructorStandingsId	raceId	constructorId	points	position	positionText	wins
1	18	1	14	1	1	1
2	18	2	8	3	3	0
3	18	3	9	2	2	0
4	18	4	5	4	4	0
5	18	5	2	5	5	0
6	18	6	1	6	6	0
7	19	1	24	1	1	1
8	19	2	19	2	2	0
9	19	3	9	4	4	0
10	19	4	6	5	5	0

Table 6: Constructors Data

constructorId	constructorRef	name	nationality	url
1	mclaren	McLaren	British	http://en.wikipedia.org/wiki/McLaren
2	bmw_sauber	BMW Sauber	German	http://en.wikipedia.org/wiki/BMW_Sauber
3	williams	Williams	British	http://en.wikipedia.org/wiki/Williams_Grand_Prix_Eng
4	renault	Renault	French	http://en.wikipedia.org/wiki/Renault_in_Formula_One
5	toro_rosso	Toro Rosso	Italian	http://en.wikipedia.org/wiki/Scuderia_Toro_Rosso
6	ferrari	Ferrari	Italian	http://en.wikipedia.org/wiki/Scuderia_Ferrari
7	toyota	Toyota	Japanese	http://en.wikipedia.org/wiki/Toyota_Racing
8	super_aguri	Super Aguri	Japanese	http://en.wikipedia.org/wiki/Super_Aguri_F1
9	red_bull	Red Bull	Austrian	http://en.wikipedia.org/wiki/Red_Bull_Racing
10	force_india	Force India	Indian	http://en.wikipedia.org/wiki/Racing_Point_Force_India

Table 7: Driver Standing Data

driverStandingsId	raceId	driverId	points	positionText	wins	NA
1	18	1	10	1	1	1
2	18	2	8	2	2	0
3	18	3	6	3	3	0
4	18	4	5	4	4	0
5	18	5	4	5	5	0
6	18	6	3	6	6	0
7	18	7	2	7	7	0
8	18	8	1	8	8	0
9	19	1	14	1	1	1
10	19	2	11	3	3	0

Table 8: Driver Data

driverId	driverRef	number	code	forename	surname	dob	nationality	url
1	hamilton	44	HAM	Lewis	Hamilton	1985-01-07	British	http://en.wikipedia.org/wi
2	heidfeld	\N	HEI	Nick	Heidfeld	1977-05-10	German	http://en.wikipedia.org/wi
3	rosberg	6	ROS	Nico	Rosberg	1985-06-27	German	http://en.wikipedia.org/wi
4	alonso	14	ALO	Fernando	Alonso	1981-07-29	Spanish	http://en.wikipedia.org/wi
5	kovalainen	\N	KOV	Heikki	Kovalainen	1981-10-19	Finnish	http://en.wikipedia.org/wi
6	nakajima	\N	NAK	Kazuki	Nakajima	1985-01-11	Japanese	http://en.wikipedia.org/wi
7	bourdais	\N	BOU	Sébastien	Bourdais	1979-02-28	French	http://en.wikipedia.org/wi
8	raikkonen	7	RAI	Kimi	Räikkönen	1979-10-17	Finnish	http://en.wikipedia.org/wi
9	kubica	\N	KUB	Robert	Kubica	1984-12-07	Polish	http://en.wikipedia.org/wi
10	glock	\N	GLO	Timo	Glock	1982-03-18	German	http://en.wikipedia.org/wi

Table 9: Lap Times Data

raceId	driverId	lap	position	time	milliseconds
841	20	1	1	1:38.109	98109
841	20	2	1	1:33.006	93006
841	20	3	1	1:32.713	92713
841	20	4	1	1:32.803	92803
841	20	5	1	1:32.342	92342
841	20	6	1	1:32.605	92605
841	20	7	1	1:32.502	92502
841	20	8	1	1:32.537	92537
841	20	9	1	1:33.240	93240
841	20	10	1	1:32.572	92572

Table 10: Pit Stops Data

raceId	driverId	stop	lap	time	duration	milliseconds
841	153	1	1	17:05:23	26.898	26898
841	30	1	1	17:05:52	25.021	25021
841	17	1	11	17:20:48	23.426	23426
841	4	1	12	17:22:34	23.251	23251
841	13	1	13	17:24:10	23.842	23842
841	22	1	13	17:24:29	23.643	23643
841	20	1	14	17:25:17	22.603	22603
841	814	1	14	17:26:03	24.863	24863
841	816	1	14	17:26:50	25.259	25259
841	67	1	15	17:27:34	25.342	25342

Table 11: Qualifying Data

qualifyId	raceId	driverId	constructorId	number	position	q1	q2	q3
1	18	1	1	22	1	1:26.572	1:25.187	1:26.714
2	18	9	2	4	2	1:26.103	1:25.315	1:26.869
3	18	5	1	23	3	1:25.664	1:25.452	1:27.079
4	18	13	6	2	4	1:25.994	1:25.691	1:27.178
5	18	2	2	3	5	1:25.960	1:25.518	1:27.236
6	18	15	7	11	6	1:26.427	1:26.101	1:28.527
7	18	3	3	7	7	1:26.295	1:26.059	1:28.687
8	18	14	9	9	8	1:26.381	1:26.063	1:29.041
9	18	10	7	12	9	1:26.919	1:26.164	1:29.593
10	18	20	5	15	10	1:26.702	1:25.842	\N

Table 12: Race Data

raceId	year	round	circuitId	name	date	time	url
1	2009	1	1	Australian Grand Prix	2009-03-29	06:00:00	http://en.wikipedia.org/wiki/2009_
2	2009	2	2	Malaysian Grand Prix	2009-04-05	09:00:00	http://en.wikipedia.org/wiki/2009_
3	2009	3	17	Chinese Grand Prix	2009-04-19	07:00:00	http://en.wikipedia.org/wiki/2009_
4	2009	4	3	Bahrain Grand Prix	2009-04-26	12:00:00	http://en.wikipedia.org/wiki/2009_
5	2009	5	4	Spanish Grand Prix	2009-05-10	12:00:00	http://en.wikipedia.org/wiki/2009_
6	2009	6	6	Monaco Grand Prix	2009-05-24	12:00:00	http://en.wikipedia.org/wiki/2009_
7	2009	7	5	Turkish Grand Prix	2009-06-07	12:00:00	http://en.wikipedia.org/wiki/2009_
8	2009	8	9	British Grand Prix	2009-06-21	12:00:00	http://en.wikipedia.org/wiki/2009_
9	2009	9	20	German Grand Prix	2009-07-12	12:00:00	http://en.wikipedia.org/wiki/2009_
10	2009	10	11	Hungarian Grand Prix	2009-07-26	12:00:00	http://en.wikipedia.org/wiki/2009_

Table 13: Result Data

resultId	raceId	driverId	constructorId	number	grid	position	positionText	positionOrder	points	laps	t
1	18	1	1	22	1	1	1	1	10	58	
2	18	2	2	3	5	2	2	2	8	58	
3	18	3	3	7	7	3	3	3	6	58	
4	18	4	4	5	11	4	4	4	5	58	
5	18	5	1	23	3	5	5	5	4	58	
6	18	6	3	8	13	6	6	6	3	57	
7	18	7	5	14	17	7	7	7	2	55	
8	18	8	6	1	15	8	8	8	1	53	
9	18	9	2	4	2	\N	R	9	0	47	
10	18	10	7	12	18	\N	R	10	0	43	

Table 14: Season Data

year	url
2009	http://en.wikipedia.org/wiki/2009_Formula_One_season
2008	http://en.wikipedia.org/wiki/2008_Formula_One_season
2007	http://en.wikipedia.org/wiki/2007_Formula_One_season
2006	http://en.wikipedia.org/wiki/2006_Formula_One_season
2005	http://en.wikipedia.org/wiki/2005_Formula_One_season
2004	http://en.wikipedia.org/wiki/2004_Formula_One_season
2003	http://en.wikipedia.org/wiki/2003_Formula_One_season
2002	http://en.wikipedia.org/wiki/2002_Formula_One_season
2001	http://en.wikipedia.org/wiki/2001_Formula_One_season
2000	http://en.wikipedia.org/wiki/2000_Formula_One_season

Table 15: Status Data

statusId	status
1	Finished
2	Disqualified
3	Accident
4	Collision
5	Engine
6	Gearbox
7	Transmission
8	Clutch
9	Hydraulics
10	Electrical

8 References

Wickham et al. (2018) Allaire et al. (2015) Wickham and others (2014) Sulsters and Bekker (2018) Bhatnagar (2014) Spenke and Beilken (2000) Wickham (2016) Stewart (2018) FIA (n.d.) Davies and Masters (2016)

Allaire, JJ, Joe Cheng, Yihui Xie, Jonathan McPherson, Winston Chang, Jeff Allen, Hadley Wickham, Aron Atkins, and Rob Hyndman. 2015. "Rmarkdown: Dynamic Documents for R." *R Package Version 0.5*.

Bhatnagar, Unmukt Rajeev. 2014. "FORMULA 1 Race Car Performance Improvement by Optimization of the Aerodynamic Relationship Between the Front and Rear Wings."

Davies, Amanda, and James Masters. 2016. "Jules Bianchi: The Death That Rocked F1." *CNN*. Cable News Network. https://www.cnn.com/2016/07/14/motorsport/jules-bianchi-deat-anniversary-formula-one/index.html.

FIA. n.d. "Regulations." *Federation Internationale de L'Automobile*. https://www.fia.com/regulation/category/110.

Spenke, Michael, and Christian Beilken. 2000. "InfoZoom-Analysing Formula One Racing Results with an Interactive Data Mining and Visualisation Tool." *WIT Transactions on Information and Communication Technologies* 25. WIT Press.

Stewart, Jack. 2018. "Witness 60 Years of Formula One Race Car Evolution." *Wired*. Conde Nast. https://www.wired.com/2017/03/formula-one-f1-race-car-evolution/.

Sulsters, Claudia, and R Bekker. 2018. "Simulating Formula One Race Strategies."

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2018. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and others. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10). Foundation for Open Access Statistics: 1–23.