Final – Project Writeup and Visualization

Name: *Matthew Godbold*

Project Title: *Data Fueled Insights: Analyzing Formula 1 Through Interactive Visualizations*

I. Introduction / Background

Formula 1 has been renowned as the pinnacle of motorsports for the past 70 years. Its single seat, open-wheel, open cockpit, high-tech race cars are a marvel of engineering, both mechanically and aerodynamically. However, while many fans appreciate the speed and skill of the drivers, few are aware of the incredible amount of data collected from the cars, tracks, drivers, and pits.

In fact, the sheer amount of data collected is a masterful conduction of data collection and correlation. By communicating this data to its audience, Formula 1 can offer fans a deeper understanding of the sport, allowing them to explore their favorite drivers, constructors, and tracks in a way incomparable to any other sport. With so much data available, it's no surprise that fans are eager to get their hands on it. By creating a data visualization dashboard, fans can explore and analyze the data, gaining valuable insights into the performance of their favorite drivers and constructors. From lap times and tire wear to fuel consumption and engine performance, the possibilities are endless. By providing fans with this level of insight, Formula 1 is not only enhancing the fan experience, but also providing valuable data that can be used to improve the performance of the cars and drivers themselves.

II. Audience

Formula 1 has a global fanbase of nearly 1.55 billion and has the highest YoY increase of any sport at nearly 30%. It also crowns the sports entertainment industry for social media engagement with an active following of 50 million users. This is made possible by the social media teams, utilizing data in an easily comprehendible way. To give fans centralized data visualizations, I am building a dashboard for interactive visualizations.

III. Dataset(s)

Source: *https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?select=status.csv*

Description:

The data set consists of thirteen CSV files (20.14 MB) containing the following features:

Circuits	
ID	*77*
Circuit Reference	*77*
Name	*77*
Location	*77*
Country	*77*
Latitude	*77*

Constructor Results	
Constructor	*12.2K*
Results ID	*12.2K*
Race ID	*12.2K*
Constructor ID	*12.2K*
Number of Points	*12.2K*

Constructor Standings	
Constructor	*12.9K*
Standings ID	*12.9K*
Race ID	*12.9K*
Constructor ID	*12.9K*
Number of Points	*12.9K*
Position Number	*12.9K*
Position Text	*12.9K*
Number of Wins	*12.9K*

Constructors	
Constructor ID	*211*
Constructor	*211*
Reference	*211*
Name	*211*
Name	*211*
Nationality	*211*
URL	*211*

Driver Standings	
Driver Standings	*33.9K*
ID	*33.9K*
Race ID	*33.9K*
Driver ID	*33.9K*
Number of Points	*33.9K*
Position Number	*33.9K*

Drivers	
Driver ID	*857*
Driver Reference	*857*
Number	*857*
Driver Number	*857*
Driver Code	*857*
First Name	*857*
Last Name	*857*
Date of Birth	*857*
Nationality	*857*
URL	*857*

Lap Times	
Race ID	*538K*
Driver ID	*538K*
Lap Number	*538K*
Position Number	*538K*
Time	*538K*
Time in ms	*538K*

Pit Stops	
Race ID	*9634*
Driver ID	*9634*
Stop Number	*9634*
Lap Number	*9634*
Time of Pit Stop	*9634*
Duration	*9634*
Duration in ms	*9634*

Qualifying	
Qualify ID	*9575*
Race ID	*9575*
Driver ID	*9575*
Constructor ID	*9575*
Car Number	*9575*
Position Number	*9575*
Q1 Time	*9575*
Q2 Time	*9575*
Q3 Time	*9575*

Races	
Race ID	*1102*
Year	*1102*
Round	*1102*
Circuit ID	*1102*
Name	*1102*
Date	*1102*
Time	*1102*
URL	*1102*

Results	
Result ID	*25.8K*
Race ID	*25.8K*
Driver ID	*25.8K*
Constructor ID	*25.8K*
Car Number	*25.8K*
Grid Position	*25.8K*
Final Position	*25.8K*
Position Text	*25.8K*
Position Order	*25.8K*
Number of Points	*25.8K*

Foreseen Issues:
At this point in the project, there are no foreseen issues. All data appears to be cleaned and prepared for visualizations.

IV. Proposed Visualizations

description of visualizations you intend to create
describe any interactivity that you intend to include

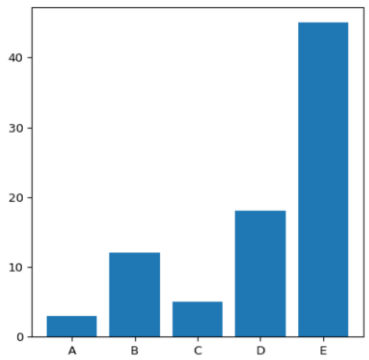
*include example images for the **types** of visualizations*
example images can come from unrelated datasets and sources
cite any images you include from external sources (if it's from a website, the URL is fine)

Formula 1 is a unique sport in that the competition is multifaceted. There are two main competition levels: the driver and the constructor. On the driver level, there are twenty drivers that compete for points (earned based on finishing position) that go towards winning the Driver's Championship. On the other hand, the ten constructors (comprised of two drivers each) compete for team points which are based on the points earned by your two drivers. For this reason, I plan to create a dashboard that is separated into two main pages, one for drivers and one for constructors.

Driver Visualizations:

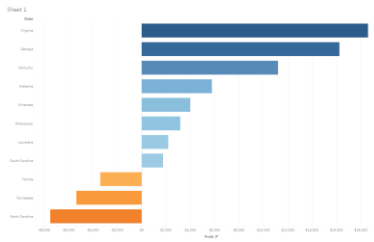
Driver Success

Description: *A bar chart comparing the driver vs. their wins and points per grand prix.*
Interactivity: *The user will be able to select the driver of their choice.*
Source: *https://www.python-graph-gallery.com/barplot/*



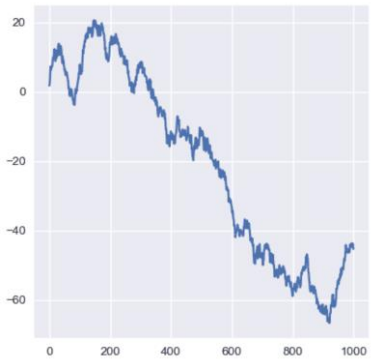
Driver Conversions

Description: *A diverging bar chart showing the drivers conversions per grand prix.*
Interactivity: *The user will be able to select the driver.*
Source:[*https://www.thedataschool.co.uk/timothy-manning/make-clean-diverging-bar-chart-tableau-tips-tableautimothy*](https://www.thedataschool.co.uk/timothy-manning/make-clean-diverging-bar-chart-tableau-tips-tableautimothy)



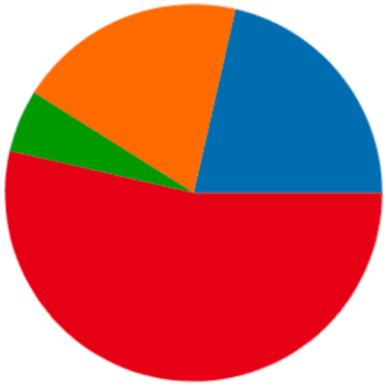
Driver Position

Description: *A line chart will be used to show the drivers lap time per lap and position per lap for each grand prix.*
Interactivity: *The user will be able to select the driver, lap time or position, and the grand prix.*
Source: [*https://www.python-graph-gallery.com/line-chart/*](https://www.python-graph-gallery.com/line-chart/)



Pit Duration

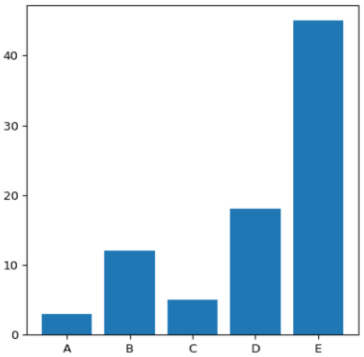
Description: *A pie chart will be used to show the time the driver spent in the pit during the total duration of the race.*
Interactivity: *The user will be able to select the driver and grand prix.*
Source: [*https://www.python-graph-gallery.com/pie-plot/*](https://www.python-graph-gallery.com/pie-plot/)



Constructor Visualizations:

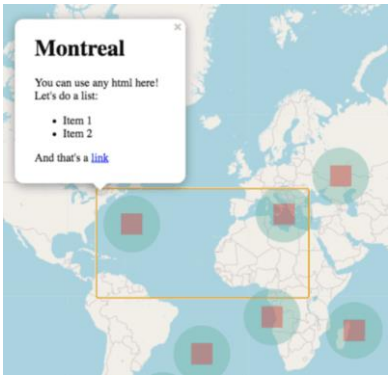
Constructor Success

Description: *A bar chart comparing the constructor wins and points vs other constructors per grand prix or all time.*
Interactivity: *The user will be able to select the constructor, grand prix or all time, and if they view wins or points.*
Source: *https://www.python-graph-gallery.com/barplot/*



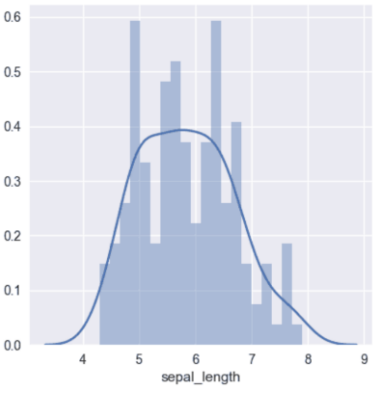
Mapping the Team

Description: *Map placing the nationality of each of its drivers.*
Interactivity: *The user will be able to select the constructor of their choice.*
Source: *https://www.python-graph-gallery.com/map/*



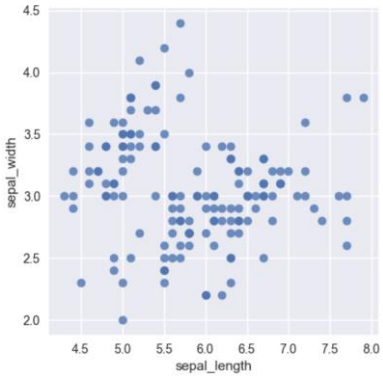
Points Garnered

Description: *A histogram showing which races garnered the most points for the team in a season or all time.*
Interactivity: *The user can select the constructor and the year or all time.*
Source: *https://www.python-graph-gallery.com/histogram/*



Driver's Over Time

Description: *Scatter plot showing the constructor's driver positions overtime.*
Interactivity: *User is able to select the constructor.*
Source: *https://www.python-graph-gallery.com/scatter-plot/*



V. Plan

Plan A:

To build the visualizations, Python will be employed along with several common packages namely: Polars, Matplotlib, Plotly, and Seaborn. Streamlit will be used to build the dashboard and all user capabilities.

Plan B:

If things don't go to plan, Tableau will be used to create the visualizations and dashboard. This is the backup because Tableau provides an easy to use, all-in-one tool for creating the application.

VI. Feedback

Instructor:

This is an interesting dataset with lots of visualizations on Kaggle. I look forward to seeing the new visualizations you create for this project!

Peer Reviewer 1:

Attached in the Appendix.

Peer Reviewer 2:

Attached in the Appendix.

VII. Changes

The comments provided by the instructor and two peer reviewers provided validation of the visualization dashboard, and only one change was suggested: "It lacks how the visualization is going to be delivered to users. In the above section it mentions how the user is going to interact with the visualization, but it does not inform the reader of the method of delivery." Based on this suggestion, I rethought the explanation of my method of delivery. Streamlit is an open-source Python library that enables developers to rapidly create interactive, web-based data applications through an intuitive and user-friendly interface. These applications can be effortlessly published to end users by deploying the app directly on social media or sharing it through their community forum, offering an interactive way to explore, analyze, and visualize data in a web-based environment. In section IX, further details are provided to access the visualization application.

VIII. Visualization Summary

"Data Fueled Insights: Analyzing Formula 1 Through Interactive Visualizations" presents a captivating exploration into the world of Formula 1, showcasing the performance of drivers and constructors over time. This data-driven story reveals the fascinating interplay between drivers, constructors, and their success across various Grand Prix races, uncovering patterns and trends that offer a deeper understanding of the sport.

Accessible through a user-friendly dashboard, the visualization is designed for a general audience with an interest in Formula 1. To begin your journey, simply navigate to the provided URL. Once you have accessed the visualization, you will be greeted with a captivating introduction to the world of Formula 1 and the significance of the data presented.

The dashboard is organized into separate sections for drivers and constructors, each containing interactive charts tailored to their respective data. Start by exploring the driver section, where you can compare driver performance in terms of wins and points using bar charts. Delve deeper into the data by examining diverging bar charts displaying driver conversions per Grand Prix, and analyze pit stop durations with an insightful pie chart.

Moving on to the constructor section, you will find bar charts comparing constructor wins and points, an engaging map showcasing the nationalities of each team's drivers, and a histogram highlighting the races that garnered the most points for each constructor.

To engage with the visualizations, simply interact with the various options available for each chart, such as selecting drivers, constructors, or datasets. This allows you to customize your experience and focus on the aspects of Formula 1 that interest you the most. Allowing discovery of hidden patterns, fascinating trends, and intriguing stories that bring the world of motorsports to life.

IX. Visualization Access

To access the data visualization application, follow the link below:
https://godboldm-csce567-final-app-7cf9zg.streamlit.app/

X. Challenges

In the process of building the Streamlit application, several technical challenges emerged that required modifications to the initial proposal. Initially, creating the application proved more difficult than anticipated, with the integration and interoperability of various datasets posing a significant hurdle. Furthermore, the ambitious scope of presenting data for multiple drivers, constructors, years, and grand prix events added complexity to the project. To address these issues and ensure a polished and functional final product, the decision was made to simplify the application by removing one visualization from both the driver and constructor sections. Additionally, the scope was narrowed to exclusively cover the 2022 season (the last completed season), streamlining the data handling process. These adjustments allowed for a seamless user experience, providing fans with interactive and engaging visualizations that deepened their understanding of the sport.

XI. Design Decisions

The layout of the application is divided into two distinct sections: a main area that houses information, visualizations, and interactive elements, and a sidebar featuring an application navigation menu. The navigation menu is organized into five separate pages, beginning with an

introductory landing page that welcomes users to the visualization dashboard and offers an overview of Formula 1.
The subsequent driver and constructor pages present the core visualizations, while the final two sections grant users access to the datasets employed in these visualizations, along with the associated midterm and final reports. This streamlined layout is designed to guide users seamlessly through the application while simultaneously enabling them to uncover new insights and patterns in the data that might have previously gone unnoticed.

Driver 1:
In creating the visualization, I opted for a bar chart to compare the drivers' wins and points per Grand Prix, as it effectively demonstrates the relative performance of each driver. Bar charts are an intuitive and straightforward way to display comparisons, and they enable viewers to quickly discern patterns and differences among drivers.
Other possible chart types, such as line charts or scatter plots, were considered, but the bar chart emerged as the most suitable choice for this data. Line charts, for example, are better suited for illustrating trends over time, while scatter plots are ideal for showing the correlation between two variables. In this case, the bar chart provided the clearest representation of the drivers' performance in terms of wins and points.
For the color scheme, an appropriate palette was selected to ensure visual appeal and easy interpretation. Colors were chosen to minimize visual clutter and allow users to focus on the data itself.
Regarding interactivity, I decided to give users the option to select the driver of their choice. This feature was incorporated to offer users a personalized experience and enable them to focus on the drivers they find most interesting or relevant. Alternative interactivity options could have included the ability to filter data by year, race, or team, but it was ultimately determined that the driver selection would provide the most valuable insights for the target audience.

Driver 2:
For the visualization showcasing driver conversions per Grand Prix, a diverging bar chart was chosen as the most appropriate type of chart. This design choice was made for several reasons. Diverging bar charts are excellent for displaying data that has a central reference point or for comparing differences between two groups. In this case, the chart effectively conveys the laps gained or lost by the driver in comparison to their starting position.
Alternative chart types, such as line charts or area charts, were considered but ultimately dismissed. While these options could display the data, the diverging bar chart provides a clearer and more intuitive representation of the driver's progress or setbacks during each Grand Prix.
The color scheme for this visualization was chosen to maintain consistency with the other visualizations in the dashboard, ensuring a cohesive user experience.
As for interactivity, users can select the driver they want to analyze. This feature empowers users to focus on the driver they are most interested in and explore their performance throughout different Grand Prix events. While other interactive options, such as filtering by year or race, could have been implemented, the driver selection was deemed the most relevant and engaging choice for the target audience.

Driver 3:

In the creation of the visualization displaying the time drivers spent in the pit during the total duration of the race, a pie chart was chosen as the ideal chart type. This decision was made because pie charts are particularly effective in showing the relative proportions of different categories within a whole. In this instance, the pie chart allows users to easily compare the pit times of all drivers during a specific race, emphasizing how much each driver contributed to the total pit time.
Other chart types, such as bar charts or stacked bar charts, were also considered. However, these alternatives were not as efficient in demonstrating the relative proportions of pit times as a pie chart. The pie chart's design allows for a clear visual representation of each driver's share of the total pit time, enabling users to quickly discern patterns and differences.
For interactivity, users can choose both the driver and the Grand Prix they wish to examine. This feature allows for a tailored analysis of each driver's performance during a specific race. By providing users with the ability to select different races, visualization encourages exploration and promotes the discovery of insights related to pit strategies and race outcomes. While additional interactive options could have been implemented, such as filtering by team or season, the driver and Grand Prix selection were deemed the most engaging and relevant choices for the target audience.

Constructor 1:
In designing the visualization for constructors, similar design choices were made as for the drivers, with a bar chart comparing constructor wins and points against other constructors per Grand Prix or all time. As previously mentioned, bar charts are an effective way to display comparisons, allowing viewers to quickly discern patterns and differences among constructors.
While alternative chart types were considered, the bar chart was once again chosen for its ability to best convey the data at hand. Pie charts, for instance, could be used to show the proportion of wins or points for each constructor, but they would not effectively illustrate the direct comparison between constructors as the bar chart does.
Regarding interactivity, users can select the constructor and choose whether to view wins or points. This level of interactivity empowers users to explore the data based on their interests and gain insights specific to the constructor they find most relevant. Alternative interactivity options, such as filtering data by year or race, were considered, but it was determined that constructor selection and the choice between wins and points provided the most valuable and engaging experience for the target audience.

Constructor 2:
In creating the visualization that highlights the nationality of each constructor's drivers, a map was selected as the primary chart type to effectively represent geographical information. This choice allows users to quickly understand the international reach of their chosen constructor and observe patterns in driver recruitment.
While other visualization options, such as tables or bar charts, were explored, a map was deemed the most effective way to display location-based data. Users can easily grasp the global distribution of drivers without needing to cross-reference countries and nationalities, resulting in an intuitive representation of geographical information.
For interactivity, users are given the ability to select interact with the map, offering a tailored analysis and promoting further exploration. Although alternative interactive options, like filtering by driver nationality or season, could have been included, the constructor selection was found to

be the most relevant choice for the intended audience. This interactive feature empowers users to uncover insights about their favorite constructors' global presence and recruitment strategies, deviating from the interactivity approaches used in previous visualizations.

Constructor 3:

When designing the visualization that illustrates the points garnered by constructors during races in a season or all time, a histogram was chosen as the ideal chart type. This decision was based on the histogram's ability to effectively display the distribution of points across races, making it easier for users to identify patterns and trends in constructor performance.

Other chart types, such as line charts or bar charts, were considered, but the histogram was found to be the most suitable for showcasing the frequency of races with specific point ranges. It provides a clear, concise representation of the data, enabling users to quickly grasp the relationship between the number of points earned and the races in which they were earned.

For interactivity, users are given the option to select their constructor of interest and filter the data by year or view it across all time. This feature allows users to customize their analysis and compare the performance of constructors across different seasons or throughout the entire history of Formula 1. While other interactive options, like filtering by race location or analyzing individual driver contributions, were contemplated, the constructor and time selection approach was determined to be the most relevant for the target audience.

XII. Future Work

The work completed on this project serves as a solid foundation that can be expanded upon to provide even more insightful and engaging experiences for Formula 1 enthusiasts. One possible extension is the addition of a new visualization that showcases the impact of weather conditions on race outcomes, revealing how factors like temperature, humidity, and precipitation affect driver and constructor performances. This would require collecting and integrating weather data from various grand prix events, adding a new layer of complexity and interest to the app.

Another avenue for expansion could involve incorporating historical data from previous seasons, enabling users to compare the performance trends of drivers and constructors over time. By incorporating this additional data, fans would gain a more comprehensive understanding of the sport's evolution, as well as the factors that contribute to the success or decline of their favorite teams and drivers. Moreover, incorporating machine learning algorithms could provide predictive insights into future race outcomes based on historical trends and current performance metrics. This feature would undoubtedly captivate the audience and further elevate their engagement with the sport.

XIII. Appendix

Project Proposal Feedback Form

Proposal Author: Matthew Godbold

Project Title: Data Fueled Insights: Analyzing Formula 1 Through Interactive Visualizations

Introduction / Background

Does the proposal include a 1-2 paragraph overview?
Yes

Does the background information provided make clear the topic / purpose / intent of the project? If more explanation is needed, please describe any additional information you think would be helpful.
Yes - does a great job of introducing Formula 1 to the reader.

Audience

Does the proposal identify a reasonable audience for the visualizations?
Yes. The audience mentioned is appropriate for this type of study.
Comments:

Dataset(s)

Is / Are the dataset(s) to be used in the project identified? Yes

Does the proposal include a way to access the dataset(s)? If not, is there an explanation provided for why that access is currently available?

Does the proposal include the number of records in the dataset(s)? YEs

Does the proposal list the attributes or include the number / types of attributes for the records in the dataset(s)? Yes

Are there any anticipated data quality issues? If so, does the proposal include a plan for handling those issues?
Yes, It does.
Comments:

Proposed Visualization

Does the proposal describe the visualizations that will be created? Yes

Does the proposal describe interactivity that will be included in the visualizations? Yes

Does the proposal provide example images for the types of visualizations that will be included?
Yes

Comments: I think the visualizations mentioned are a great route for this study.

Plan

Does the proposal describe the tools that will be used to complete the project?
Yes

Does the proposed plan include a method for delivering the visualization to its viewers?
Yes

Provide your feedback regarding the proposal's ambitious Plan A.
Plan A is very doable and Python is a great tool in and of itself - and its a great idea to use the packages available For visualizations.

Provide your feedback regarding the proposal's fallback Plan B.
Tableau is also an awesome tool and its very easy resource to fall back on.

Provide any additional comments you have regarding the proposal below.
I liked that the author was so meticulous about explaining each visualization

Project Proposal Feedback Form

Proposal Author: Matthew Godbold

Project Title: Data Fueled Insights: Analyzing Formula 1 Through Interactive Visuaizations

Introduction / Background

Does the proposal include a 1-2 paragraph overview?

Does the background information provided make clear the topic / purpose / intent of the project? If more explanation is needed, please describe any additional information you think would be helpful.

Comments: Yes the proposal include two paragraph overview. I think the information is conveyed in right sequence. What, why, and how are addressed correctly making the reader know the purpose of visualizing the data and how it will benefit the formula 1 fans.

Audience

Does the proposal identify a reasonable audience for the visualizations?

Comments: Not only fans and social media team I think it will also help the team itself to easily find the factors where they are lacking in different fields if there's any and gain competitive edge.

Dataset(s)

Is / Are the dataset(s) to be used in the project identified?

Does the proposal include a way to access the dataset(s)? If not, is there an explanation provided for why that access is not currently available?

Does the proposal include the number of records in the dataset(s)?

Does the proposal list the attributes or include the number / types of attributes for the records in the dataset(s)?

Are there any anticipated data quality issues? If so, does the proposal include a plan for handling those issues?

Comments: Yes the dataset is identified and mentioned with the types of attributes going to be used for visualization in proposal which made easy to find if it has any quality issues. As per the proposal the data seems cleaned and would not need any more work on it.

Proposed Visualization

Does the proposal describe the visualizations that will be created?

Does the proposal describe interactivity that will be included in the visualizations?

Does the proposal provide example images for the types of visualizations that will be included?

Comments: The proposal describes and provides the example of visualizations to be implemented while citing and providing the interactivity.

Plan

Does the proposal describe the tools that will be used to complete the project?

Does the proposed plan include a method for delivering the visualization to its viewers?

Provide your feedback regarding the proposal's ambitious Plan A.

Provide your feedback regarding the proposal's fallback Plan B.

The proposal describes all the tools to be used in Plan A and B. Though it lacks how the visualization is going to be delivered to users. In the above section it mentions how the user is going to interact with the visualization but it does not inform the reader the method of delivery such as a report or powerpoint etc. I think plan B backs up properly if plan A does not take place as it would be efficient and straightforward to use Tableau to create the intended visualization.

Provide any additional comments you have regarding the proposal below.