

Abschlussbericht

Grundkompetenz Datenvisualisierung

Herbstsemester 2024

Autor: Adrian Meier

Fachexperten: Prof. Dr. Arzu Çöltekin
Milena Rudig

Ort, Datum: Brugg, 06.01.2025

Inhalt

1. Einleitung.....	2
1.1 Datengrundlage	2
1.2 Visualisierungen	2
1.3 Ziel des Berichts.....	2
2. LE 1 – Grundlagen der Visualisierung und Diagrammtypen.....	3
3. LE 2 – Visuelle Wahrnehmung.....	5
4. LE 3 – Designprinzipien vs. Daten.....	7
5. LE 4 – Grammatik der Grafiktools.....	9
6. LE 5 – Auswertung.....	11
7. Quellenverzeichnis	13

1. Einleitung

1.1 Datengrundlage

Als Datengrundlage für meine Abgabe nehme ich die Daten der Formel 1 Seasons, welche im Zeitraum von 1950 bis ins Jahr 2024 zur Verfügung stehen und viele unterschiedliche Daten beinhalten. Die Daten sind von Kaggle und unter folgendem Link ([Formula 1 World Championship \(1950 - 2024\)](#)) abrufbar (Vopani, 2023)

1.2 Visualisierungen

Die Visualisierungen der Daten wurden in einem Jupyter Notebook in Python mithilfe von Pandas (zum Einlesen), Matplotlib sowie Seaborn (für Grafiken) erstellt. Das Notebook sowie die Daten sind in einem Repository auf GitHub abgelegt und können unter [adrianmeier2002/GDV](#) abgerufen werden. Die Plots für diesen Bericht sind alle im Notebook «GDV Plots» zu finden. Da im Bericht selbst der Platz jeweils begrenzt ist, sind die Visualisierungen teilweise sehr klein, für eine vergrößerte Darstellung wird deshalb auf das Notebook verwiesen.

1.3 Ziel des Berichts

Ziel dieser Arbeit ist es, die fünf Lernergebnisse aus dem Modul «Grundlagen der Datenvisualisierung» zu erarbeiten. Dabei soll die Theorie hinter den einzelnen Lernergebnissen erklärt werden, sowie eine praktische Anwendung mithilfe eigener Daten und eigener Visualisierungen gemacht werden. Die Abgabe selbst beinhaltet diesen Bericht, sowie das dazugehörige Notebook und die entsprechenden Daten, welche im Repository gefunden werden können.

2. LE 1 – Grundlagen der Visualisierung und Diagrammtypen

In diesem Lernergebnis stehen die verschiedenen Typen von Diagrammen und Visualisierungen im Vordergrund. Es gibt dabei viele verschiedene Dinge zu beachten, damit man die Daten mit der passenden und korrekten Grafik darstellen kann. Ein wichtiger Punkt dabei ist, welche Daten dargestellt werden. Wir wollen im ersten Beispiel die erfolgreichsten Fahrer aller Zeiten (nach gewonnen Punkten) ermitteln und visualisieren. Dazu eignet sich ein Balkendiagramm gut (365 Data Science, 2018).

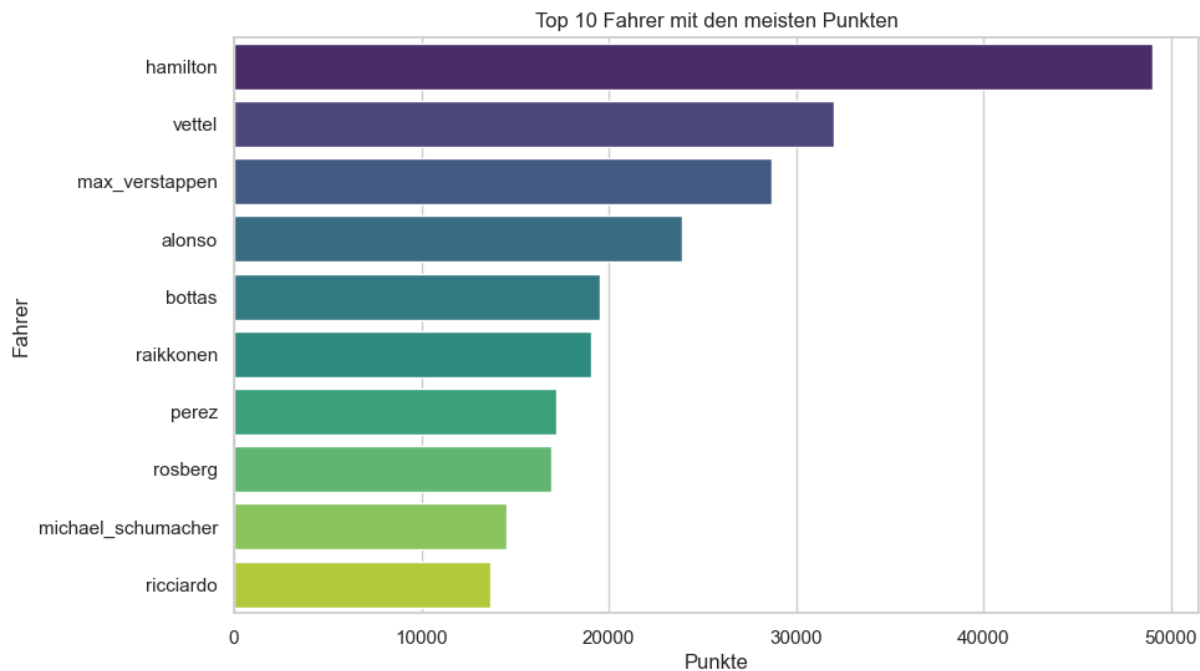


Abbildung 1: Top 10 Fahrer nach Punkten (Eigene Darstellung)

In Abbildung 1, wird schnell ersichtlich, welche Fahrer am meisten Punkte erzielt haben. Die Grafik eignet sich gut, um eine kleine Menge an Daten sauber und verständlich darzustellen.

Die Entscheidung, ein Balkendiagramm zur Visualisierung zu verwenden, wurde dadurch getroffen, dass es sich um eine numerische und eine kategoriale Variable handelt. Ein Barplot wird oft als langweilig wahrgenommen, dennoch ist es ein einfach zu verstehender Plot, welche die Daten sehr simpel und direkt darstellt. (Healy, 2018a)

Für die Darstellung der Entwicklung der einzelnen Punkte während einer Saison wird ein Linienplot gewählt. Der Hauptgrund für diese Wahl ist, dass wir Trends der einzelnen Fahrer sowie ihre Hochs und Tiefs über eine gewisse Zeit beobachten wollen, dies ist in einem Balkendiagramm nicht gut ersichtlich. In Abbildung 2 soll die Entwicklung der Punkt pro Fahrer während der Saison 2022 dargestellt werden. Aus Übersichtsgründen, wurden nur die besten 10 Fahrer der Saison im Plot berücksichtigt.

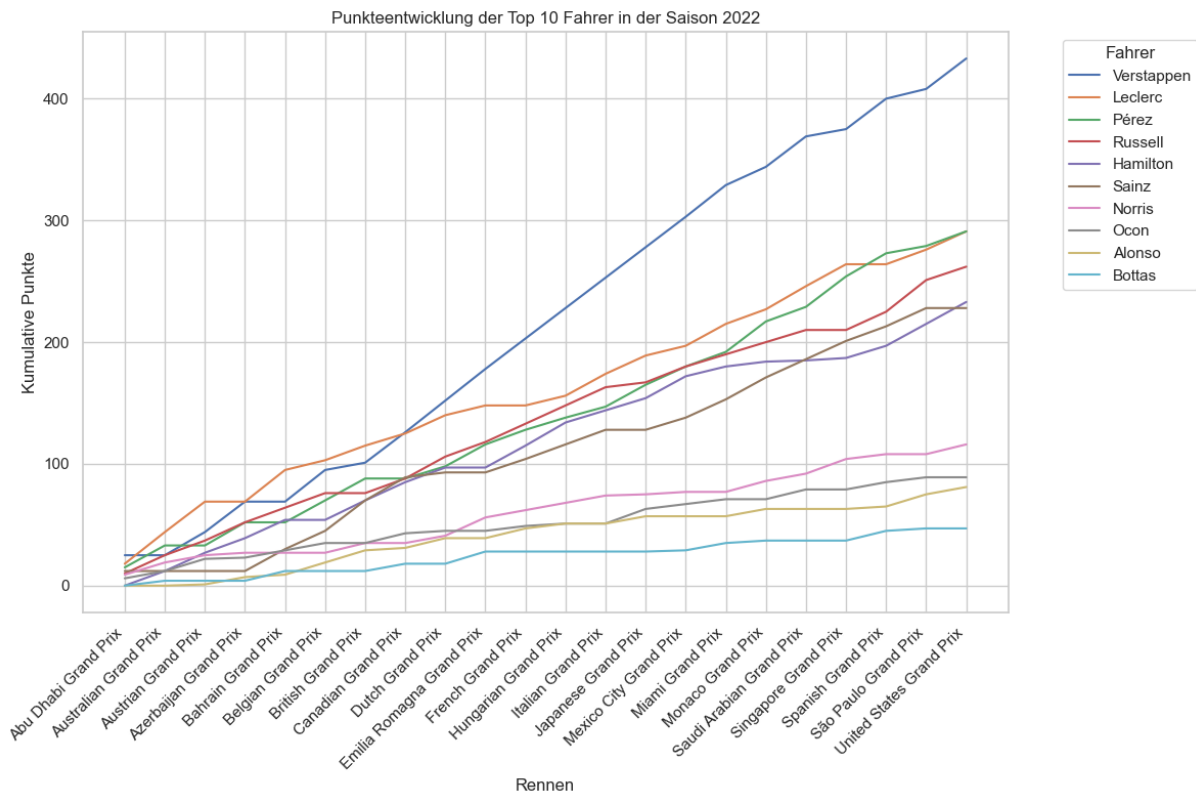


Abbildung 2: Punkteentwicklung in der Saison 2022 (Eigene Darstellung)

Im Plot ist die Entwicklung der Fahrer schön zu sehen. Man sieht, bei welchen Plätzen es eng wurde, und wie Verstappen die Saison ab dem britischen Grand Prix dominiert hat.

Die Entscheidung für diesen Linienplot kam daher, dass es eine numerische Variabel (Punkte), eine kategorische Variabel (Fahrer) und eine Zeitreihe gab. Wobei die Zeitreihe in diesem Beispiel die unterschiedlichen Rennen sind, die nacheinander während der Saison gefahren wurden.

Wenn vor allem die Trends im Vordergrund stehen, sind die einzelnen Messwerte wenig interessant und nebensächlich. Interessiert aber vor allem die Entwicklung eines Fahrers und dessen Punkte, werden die einzelnen Messwerte oft noch zusätzlich mit Punkten ergänzt (Abbildung 3). (Healy, 2018b)

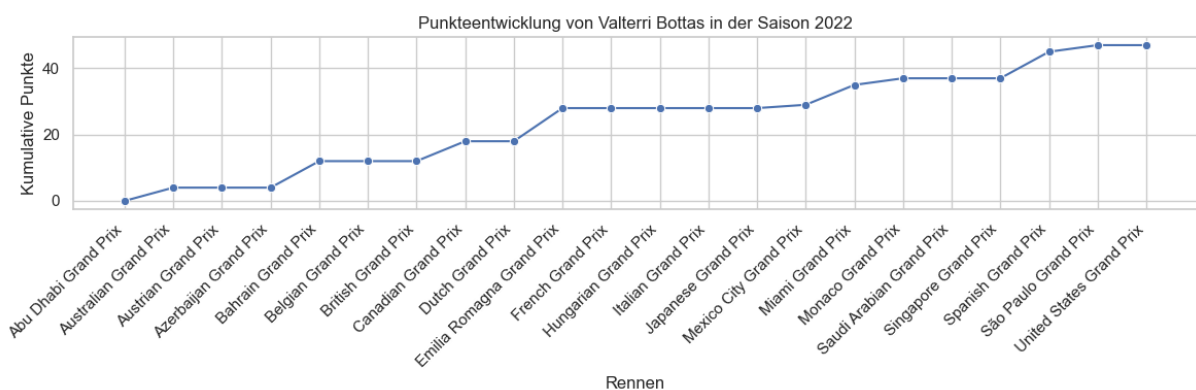


Abbildung 3: Punkteentwicklung eines Fahrers (Eigene Darstellung)

3. LE 2 – Visuelle Wahrnehmung

Um Daten visualisieren zu können ist es wichtig die visuelle Wahrnehmung von Menschen zu kennen. Dabei gibt es zwei wichtige Prinzipien: die Gestalt Psychologie und die kognitive Belastung. In der Gestalt Psychologie geht es darum, wie wir etwas wahrnehmen. Dabei wurden im Jahr 1923 sechs Gesetze durch Wertheimer definiert (Wertheimer, 1923). Die Gesetze lauten wie folgt:

Gesetz der Nähe:

Wenn Elemente geringe Abstände zueinander haben, werden diese als zusammengehörig wahrgenommen.

Gesetz der Ähnlichkeit:

Elemente, welche sich ähneln, werden einfacher als Gruppe wahrgenommen als sich nicht ähnelnde Elemente.

Gesetz der guten Gestalt:

Elemente werden besser wahrgenommen, wenn sie eine «einfache Struktur» aufweisen.

Gesetz der guten Fortsetzung:

Bei Kreuzungen von Linien wird angenommen, dass die Linien ihrem Verlauf folgen und keinen Knicks haben.

Gesetz der Geschlossenheit:

Es werden grundsätzlich Strukturen wahrgenommen, welche als geschlossen wirken.

Gesetz des gemeinsamen Schicksals:

Elemente, welche sich in die gleiche Richtung bewegen, werden als zusammengehörig angesehen.

Diese Gesetze sind wichtig zu beachten, wenn man eine Grafik erstellt und Zusammenhänge visualisieren will. Neben den oben genannten Gesetzen ist auch die kognitive Belastung, also wie viel Inhalt eine Grafik enthält, ein wichtiger Faktor, wenn es darum geht, eine gute Grafik zu erstellen. Dabei gibt es kein richtig oder falsch bei der Auswahl. Vielmehr kommt es auf das Publikum an: Wer soll diese Visualisierung verstehen? Sind es Experten oder Laien auf dem Gebiet? Und wie viel Zeit haben diese Personen, um sich die Visualisierung anzusehen? Ein wichtiger Punkt hierbei ist, das Publikum nicht mit zu viel Informationen zu überschütten aber die Visualisierung trotzdem so interessant wie möglich zu gestalten. (Sherry Seethaler, 2017)

In Abbildung 4 soll das Titelrennen in der Saison 2022 noch einmal dargestellt werden. Diesmal sind aber nur die beiden besten Teams (in dieser Saison Redbull und Ferrari) dargestellt. Zudem sind die Punkte der Teams, welche sich aus der Kumulierung der jeweiligen Fahrer ergeben, ebenfalls dargestellt.

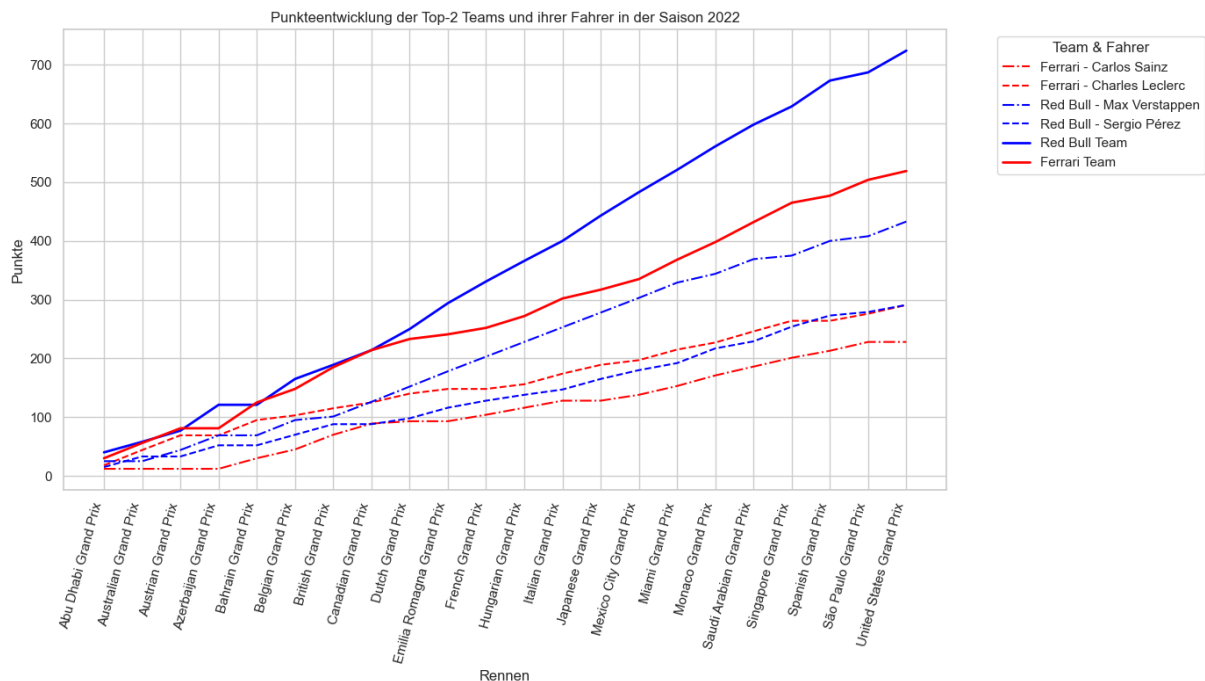


Abbildung 4: Top 2 Teams im Vergleich (Eigene Darstellung)

In diesem Beispiel ist gut zu sehen, wie sich die Punkte der Gestalt Prinzipien in der Realität umsetzen lassen, angefangen bei den Strich- beziehungsweise den Strichpunktlinien. Hier ist für das menschliche Auge klar, dass diese zusammenhängende Linien darstellen, obwohl sie lückenhaft ist. Diese Beobachtung ist ein Beispiel für das Gesetz der Geschlossenheit, der Mensch ergänzt selbständig die fehlenden Teile der Linie.

Mithilfe des Gesetzes der Ähnlichkeit können im Plot auch einfach Kategorien unterscheiden, in diesem Beispiel die zwei Teams. Während Rot die Farbe von Ferrari und ihren beiden Fahrern ist, ist Redbull mit blauen Linien gekennzeichnet. Durch die Farben wird klar, welche Linien in Bezug auf die Teams zusammengehören. Zusätzlich sind die Linien der beiden Teams als durchgehende Linie gekennzeichnet. Dies ist wieder eine Ähnlichkeit, somit kann dargestellt werden, bei welchen Linien es sich um die Teams handelt und welche zu den Fahrern gehören.

Auch das Gesetz der Fortsetzung ist zu beobachten. Auch wenn sich die Linien schneiden, ist für den Betrachter klar, wo sie nach der Kreuzung fortlaufen. Dabei wird das Ganze enorm von den verschiedenen Farben und Linientypen unterstützt, welche dabei helfen, die unterschiedlichen Linien auseinanderzuhalten. (Todorovic, 2008)

Die Visualisierung stellt nur die zwei Teams dar, welche die meisten Punkte am Ende der Saison hatten. Dabei wäre es sicherlich spannend, auch die anderen Teams zu betrachten. Hier kommt die Komplexität ins Spiel. Je nach Publikum hätte man auch noch weitere Teams und Informationen in den Plot einbauen können. Pro Team, welches aber hinzukommt, wird der Plot voller und komplizierter. Dies kann zum genauen Analysieren durch Experten, welche genügend Zeit haben, wichtig sein. Doch für die grosse Mehrheit des Publikums wäre eine solche Grafik wohl eher verwirrend. (Sibinga et al., 2021)

4. LE 3 – Designprinzipien vs. Daten

Um eine Visualisierung zu machen, braucht es zuerst eine Datengrundlage. Bevor die Daten dargestellt werden, müssen sie oft vorbearbeitet werden. Dies geschieht unter dem Begriff «Data Wrangling». Dabei werden die gesammelten Daten so aufgearbeitet, dass sie im Anschluss lesbar und verständlich sind. Erst nachdem die Daten in der korrekten Form sind, können diese sinnvoll visualisiert werden. (Wickham & Grolemund, 2017)

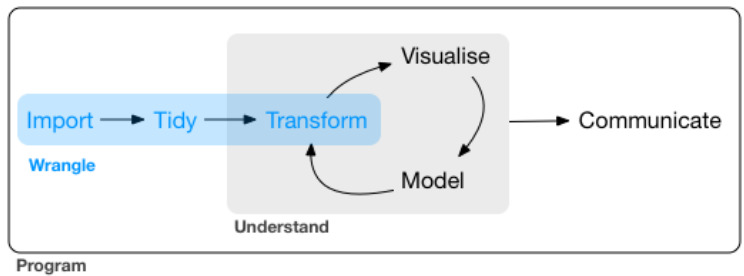


Abbildung 5: Prozess der Datenaufarbeitung (Wickham & Grolemund, 2017)

Dabei ist es wichtig zu wissen, welche Daten man überhaupt darstellt. Je nach Kategorie des Datentyps gibt es unterschiedliche Arten, diese zu visualisieren. Generell können folgende Typen unterschieden werden (Frost, 2018):

Qualitative Daten (nominal):

Bei qualitativ nominalen Daten geht es um Daten, welche als eine Kategorie angesehen werden können und dabei keine Ordnung in den Daten haben. Als Beispiel im F1 Datensatz könnte man die Herkunftsländer der Fahrer betrachten. Diese können selbst nicht logisch geordnet werden. Man müsste eine zweite Variabel einfügen, um eine Rangordnung herstellen zu können (Zum Beispiel die Länder nach Anzahl der Fahrer ordnen). Im Kuchendiagramm (Abbildung 6) sind die Länder mit dem prozentualen Anteil der Fahrer dargestellt. Es würde keinen Unterschied machen, wenn die Nationen anders angeordnet wären.

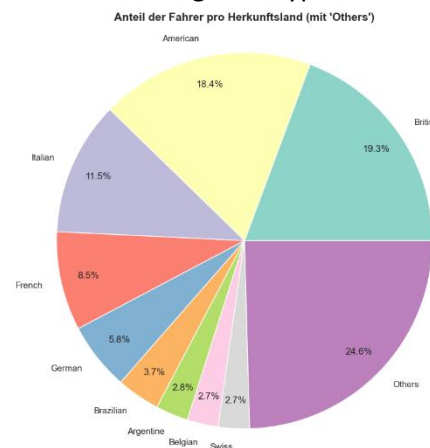


Abbildung 6: Kuchendiagramm der Fahrernationalität (Eigene Darstellung)

Qualitative Daten (ordinal):

Auch hier geht es um klare Kategorien. Dabei sind diese aber in einer gewissen Rangordnung. Wenn dies auf unseren Datensatz angewendet wird, wäre ein Beispiel dazu, einen Fahrer in Kategorien zu unterteilen, Top 10 Ergebnis (Platz 4-10), Podium (Platz 2 oder 3) oder Rennsieg (Abbildung 7). Dabei ist klar, dass ein Rennsieg höher steht als ein Podium und ein Podium besser ist als eine Top 10 Klassierung. Dies kann dann auch gut in einem Balkendiagramm dargestellt werden. Dabei sind die Balken in der Reihenfolge geordnet von «Rennsieg» nach «Ausserhalb der Top 10».

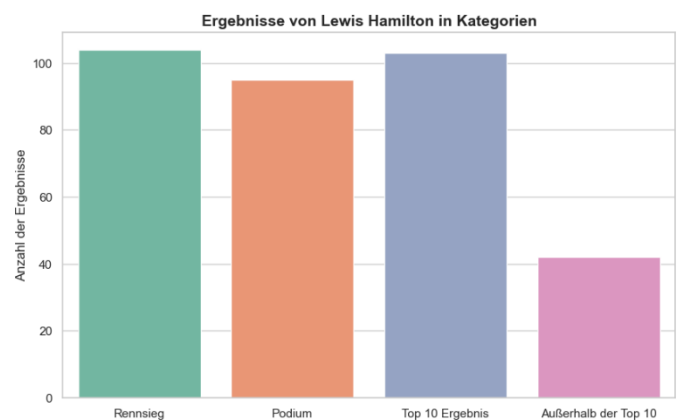


Abbildung 7: Balkendiagramm der Ergebnisse von Hamilton (Eigene Darstellung)

Quantitative Daten (diskret):

Bei den quantitativ diskreten Daten handelt es sich um durch Zahlen oder Metriken ausdrückbare Abstände in den Daten. Dabei steht das diskret dafür, dass die Zahlen nur einer endlichen Menge entsprechen und nicht weiter teilbar sind. Bei den Formel 1 Daten wären das als Beispiel die Platzierungen eines Fahrers. Diese sind klar definiert, es gibt keinen Platz 2,5 sondern nur einen zweiten oder dritten Platz. Auch hier lässt sich die vorherige Darstellung der Rennergebnisse von Lewis Hamilton gut mithilfe eines Barplots darstellen (Abbildung 8), wobei die X-Achse die diskrete Variable der Platzierungen darstellt und die Y-Achse die Anzahl aufzeigt, wie oft Hamilton diese während seiner Karriere erreicht hat.

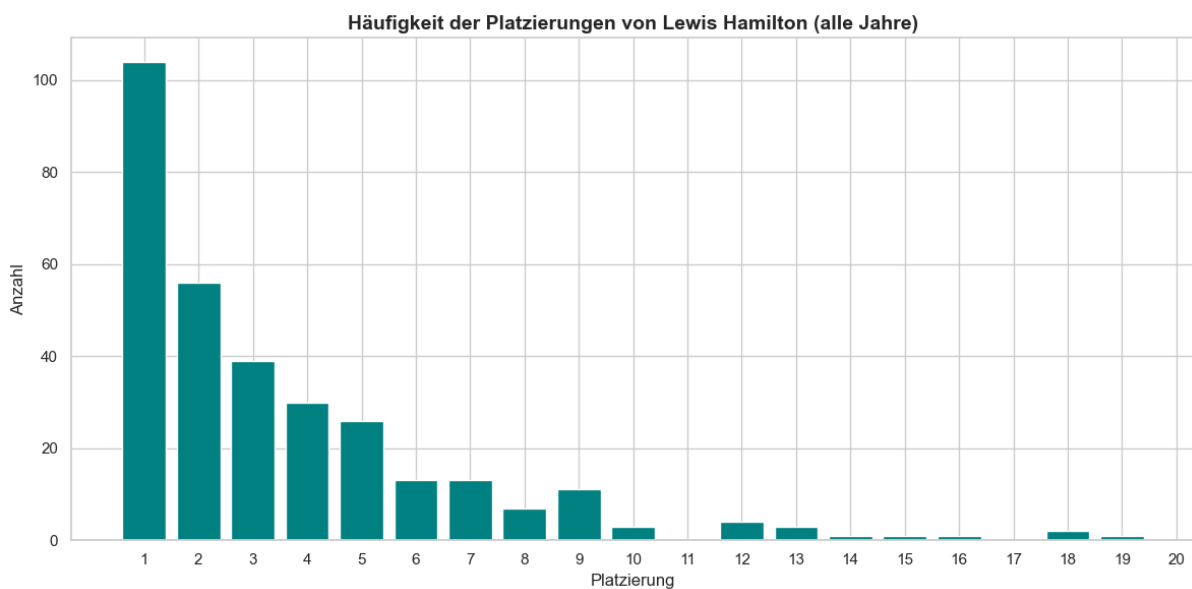


Abbildung 8: Barplot mit der Platzierung von Hamilton (Eigene Darstellung)

Quantitative Daten (stetig):

Diese Daten lassen sich unendlich genau ins Detail erforschen. Für eine Visualisierung werden sie aber oft in Bins (Intervalle) zusammengefasst, um eine saubere Darstellung machen zu können. Ein gutes Beispiel hierfür sind die Boxenstopzeiten während einer Saison. Diese sind theoretisch unendlich lange und genau, werden aber natürlich gerundet. Je nachdem, was das Ziel der Darstellung ist, können alle Punkte einzeln dargestellt werden oder wenn man nur wissen will, wie die Häufigkeit der Zeiten ist, kann es sinnvoll sein, diese zu gruppieren (In Abbildung 9 in 0,5 Sekunden Schritten). Bei den Boxenstopzeiten lässt sich die Häufigkeit mithilfe von Bins gut in einem Histogramm darstellen. Man sieht sehr schnell, dass sich die meisten Zeiten zwischen 20 und 30 Sekunden befinden.

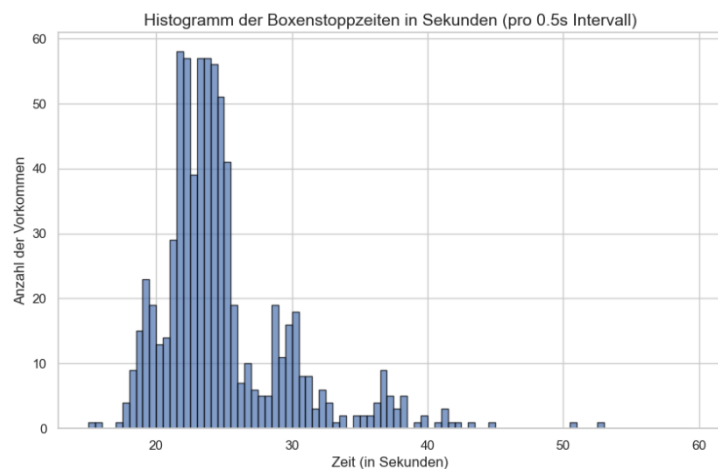


Abbildung 9: Histogramm der Boxenstops (Eigene Darstellung)

5. LE 4 – Grammatik der Grafiktools

Daten sind meistens multidimensional, dies ist auf Computern schwierig darzustellen, da ein Bildschirm eine Limitierung auf einen zweidimensionalen Raum hat. Diese multidimensionalen Räume darzustellen, ist enorm wichtig, um Zusammenhänge erkennen zu können, welche durch das reine betrachten von statistischen Kennwerten wie dem Mittelwert oder der Standardabweichung nicht dargestellt werden. Gerade, wenn man die Daten in einem späteren Verlauf weiterverarbeitet, zum Beispiel mithilfe von Machine Learning, sind Visualisierungen unabdingbar. Mithilfe von Programmen und der richtigen Anwendung von gewissen Komponenten kann man auf einem zweidimensionalen Bildschirm auch höherdimensionale Daten verständlich darstellen. Diese Komponenten kann man unterteilen in: (Sarkar, 2018)

Daten:

Der erste und wichtigste Schritt ist das Einlesen und Analysieren der Daten. Wie bereits im LE 3 angeschaut, ist es wichtig, seine Daten zu kennen und sie den verschiedenen Metriken zuzuordnen. Es ist wichtig, ein Grundverständnis für die Variablen zu haben, welche Dimensionen in den Daten vorhanden sind und wo Zusammenhänge bestehen. Wenn die Rundenzeiten und die Runde aufgezeigt werden sollen, kann dies beispielsweise mit einem Scatterplot gemacht werden. (Abbildung 10)

Ästhetik:

Welche Werte können wie dargestellt werden? Es kann sinnvoll sein, kategorialen Variablen eine Farbe pro Kategorie zuzuweisen, um eine dritte Dimension hinzuzufügen. Somit kann man zum Beispiel die Teams als Kategorie darstellen (Abbildung 11).

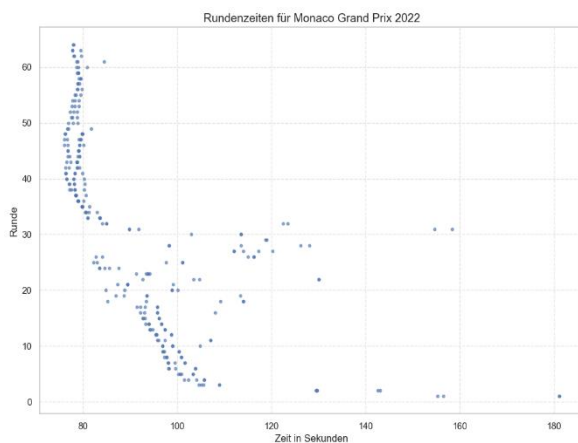


Abbildung 10: Rundenzeiten pro Runde (Eigene Darstellung)

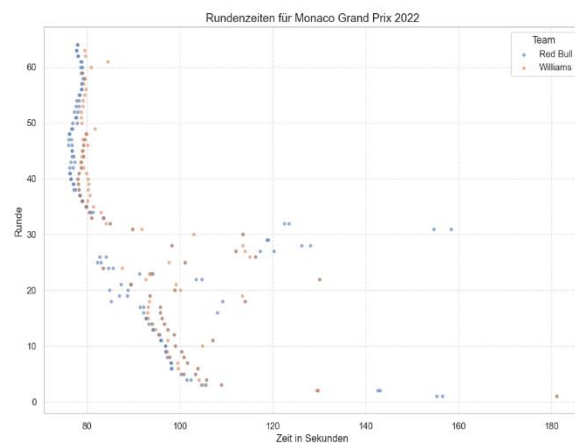


Abbildung 11: Rundenzeiten mit Teams (Eigene Darstellung)

Skalierung:

Mithilfe unterschiedlicher Größen in der Darstellung der Datenpunkte können zusätzliche Dimensionen visualisiert werden.

Geometrische Objekte:

Auch durch die Wahl der verschiedenen Objekte im Plot können Daten dargestellt werden. Hierbei wird entschieden, wie die Datenpunkte dargestellt werden, ob als Punkte, Linien, Balken usw.

Statistik:

Manchmal kann es sinnvoll sein, die Grafik durch einige statistische Ergänzungen zu unterstützen. Mit solchen Hilfestellungen können zum Beispiel Trends einzelner Gruppen dargestellt werden. Im Beispiel soll gezeigt werden, wie im Verlaufe der Runden die Zeiten immer schneller werden. Dargestellt wird dies in Abbildung 12 mithilfe einer linearen Regression, welche im Plot als rote Linie verläuft.

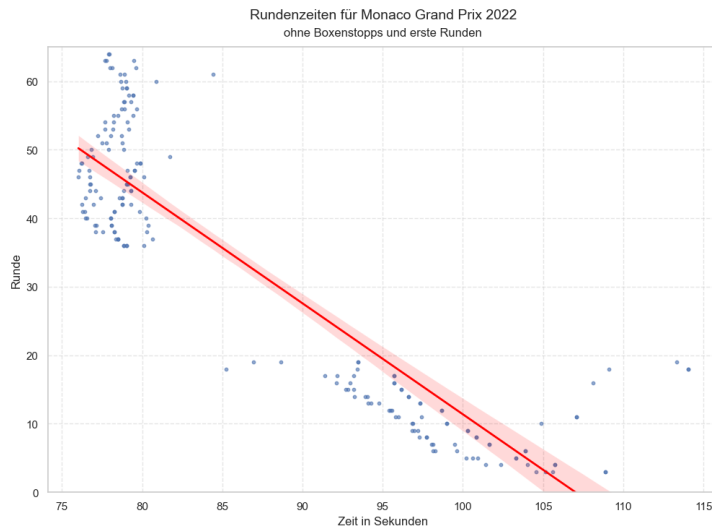


Abbildung 102: Lineare Regression auf den Rundenzeiten (Eigene Darstellung)

Facetten:

Um die Komplexität der Grafik etwas zu entwirren, können verschiedene kleinere Plots erstellt werden, welche dann nur gewisse Eigenschaften darstellen. Das Formel 1 Rennen in Monaco kann man beispielsweise in 3 Abschnitte unterteilen. Der Rennbeginn von der ersten bis circa in die 20. Runde, das Boxenstoppfenster, welches in der Mitte des Rennens stattfindet und die letzten Runden am Ende des Rennens. Mithilfe von 3 kleineren Plots kann diese Unterteilung besser veranschaulicht werden.

Koordinatensystem:

Wie soll das Koordinatensystem aussehen? Gewisse Daten lassen sich besser auf einem Polarkoordinatensystem darstellen als auf einem Kartesischen. Gerade zyklische Daten, wie beispielsweise die Temperatur über die Jahre hinweg, kann auf einem Polarkoordinatensystem oft besser dargestellt und verglichen werden. Auch hier kommt es immer auf die Intention der Visualisierung an.

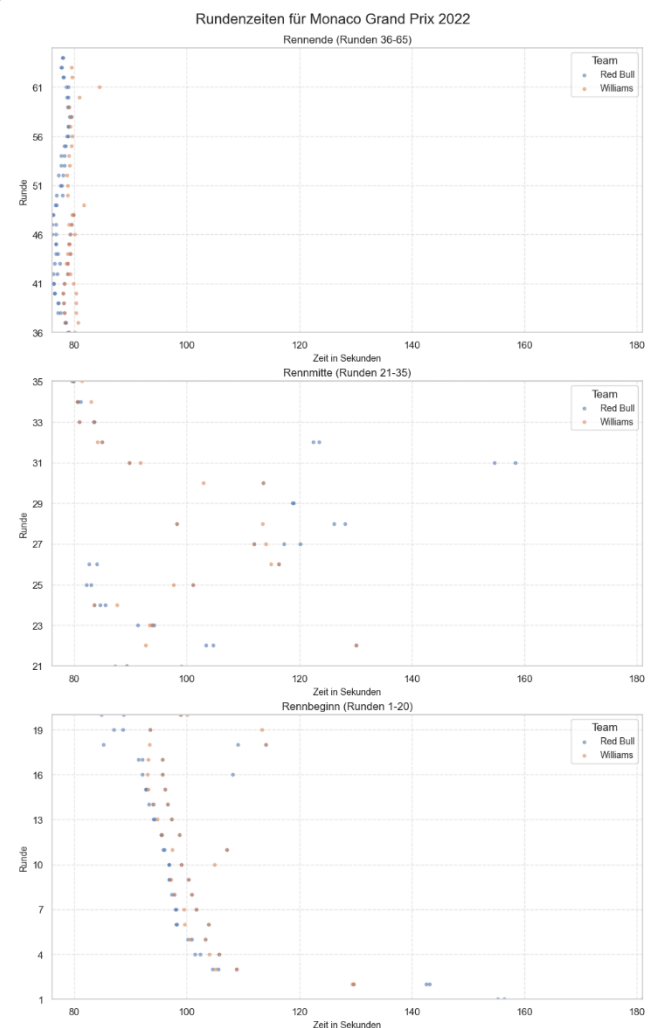


Abbildung 113: Einteilung des Rennens (Eigene Darstellung)

6. LE 5 – Auswertung

Im letzten Abschnitt geht es darum, wie unsere Visualisierungen ankommen. Am Ende ist es wichtig, dass unsere Botschaft, welche wir mit der Visualisierung ausdrücken wollen, dem Zielpublikum korrekt vermittelt wird. Um dies zu erreichen, gibt es verschiedene Arten von Auswertungen, um bereits vor der Veröffentlichung zu überprüfen, dass dies gelingt.

Eine Art dies zu prüfen, sind Usability-Tests. Dabei geht es darum, einer kleinen Gruppe von Probanden, welche dem Zielpublikum ähnlich ist. Aufgaben zu stellen und im Anschluss zu beurteilen, ob diese korrekt, einfach und schnell zu lösen waren. Dabei ist diese Testmethode sehr allgemein, sie kann vielseitig verwendet werden und ist nicht auf die Auswertungen von Visualisierungen beschränkt. Man könnte damit auch prüfen, ob das Interface einer Webseite für einen Shop verständlich ist oder ob ein Formular klar und verständlich ausgefüllt werden kann. Dabei gibt es drei wichtige Elemente (Moran, 2019):

Moderator:

Die Aufgabe der moderierenden Person ist es, die Probanden durch den Testprozess zu begleiten. Sie oder er stellt die Aufgaben, beobachtet den Probanden beim Ausführen dieser Aufgaben, beantwortet Fragen und stellt selbst Fragen an die Probanden. Ziel ist es, möglichst gute Daten zu erhalten, ohne dabei die Probanden zu beeinflussen.

Proband:

Die Aufgabe der Probanden ist es, die Aufgaben, welche von dem Moderator/ der Moderatorin gestellt werden, auszuführen und eine Rückmeldung zu geben. Dabei ist es wichtig, dass die Probanden selbst aus der Zielgruppe kommen, in dem das Produkt später angewendet werden soll. Oftmals kann es hilfreich sein, dass die Probanden ihre Gedankengänge und Aktionen während des Tests laut äussern, um noch mehr Feedback zu generieren.

Aufgabe:

Die Aufgabe, welche von der moderierenden Person gestellt wird und von den Probanden ausgeführt werden muss, soll klar und eindeutig gestellt werden und realitätsnah sein. Wie tiefgehend die jeweilige Fragestellung ist, hängt von den Tests ab. Wichtig ist, dass die Probanden die Aufgabe richtig verstanden haben, dies sollte von der moderierenden Person immer überprüft werden.

Bei den Usability Tests gibt es verschiedene Arten, wie die Tests durchgeführt werden können. Alle haben dabei Vor- sowie Nachteile und müssen dem Produkt und dem Ziel angepasst werden. Folgende Unterteilungen können gemacht werden (Moran, 2019):

Qualitativ oder quantitativ:

- Qualitativ:
Beim qualitativen Testen ist das Ziel herauszufinden, wie die Personen das Produkt nutzen. Dabei soll erkannt werden, wo die Probanden noch Probleme haben und wie sich die Bedienung für sie anfühlt.
- Quantitativ:
Das quantitative Testen soll messbare Metriken generieren. Meist wird dabei die Erfolgsrate bei der Abschliessung der Aufgabe getestet und wie viel Zeit die Probanden dafür benötigt haben.

Arten der Durchführung:

- Moderiert:
Dies kann entweder vor Ort von Moderator*in zu Proband*in oder Remote stattfinden. Beim Remote befinden sich die Testpersonen und der Moderator zwar physisch nicht im selben Raum, aber haben online eine Verbindung zueinander (Skype, Zoom, Teams etc).
- Unmoderiert:
Die Testpersonen sind nicht im direkten Kontakt mit dem Moderierenden. Dies wird meistens mit Hilfe eines Online-Tests durchgeführt.

Anwendung auf Datenvisualisierungen:

Bevor ein solcher Usability Test durchgeführt werden kann, ist es notwendig zu wissen, wer die Grafik überhaupt sehen soll und was die Umstände sind. Ist es eine Präsentation an Experten, welche sich eine längere Zeit mit der Grafik beschäftigen sollen? Ist es eine Werbung, bei welcher innerhalb weniger Sekunden der ganze Inhalt klar sein muss? Auf diesen Ausgangssituationen baut anschliessend der Test auf. Es gibt verschiedene Arten eine Grafik zu testen. Beispielsweise kann mittels Eye-Tracking, also dem verfolgen des Fokus der Augen, gearbeitet werden, um herauszufinden, welche Punkte einer Grafik am meisten betrachtet wird. Im Falle einer Werbung könnte ein 5-Sekunden Test sinnvoll sein. Bei diesem wird die Reaktion der testenden Person nach einer sehr kurzen Zeit bewertet, somit kann sichergestellt werden, dass eine Abbildung ansprechend ist.

Nach einem Test sollen die Probanden ein Feedback geben. Dies kann durch offene Fragen, vorgegebene Antworten oder Skalen gemacht werden.

Das Wichtigste beim Testen von eigenen Grafiken ist immer, dass man sich bewusst ist, dass man selbst nicht der User ist und deshalb auch nicht bewerten kann, welche Aspekte einer Grafik gut aussehen und intuitiv sind. Dies wird am Ende nur durch das Publikum beurteilt (Chan, 2024).

7. Quellenverzeichnis

- 365 Data Science (Regisseur). (2018, September 4). *Which is the best chart: Selecting among 14 types of charts Part I* [Video recording]. <https://www.youtube.com/watch?v=C07k0euBpr8>
- Chan, M. (2024, Dezember 13). *Testing Visual Design: A Comprehensive Guide*. Nielsen Norman Group. <https://www.nngroup.com/articles/testing-visual-design/>
- Frost, J. (2018, Januar 30). *Guide to Data Types and How to Graph Them in Statistics*. Statistics By Jim. <http://statisticsbyjim.com/basics/data-types/>
- Healy, Y. H. and C. (2018a, Januar 1). *Barplot*. data-to-viz. <https://www.data-to-viz.com/graph/www.data-to-viz.com/caveat/barplot.html>
- Healy, Y. H. and C. (2018b, Januar 1). *Line chart*. <https://www.data-to-viz.com/graph/www.data-to-viz.com/caveat/line.html>
- Moran, K. (2019, Dezember 1). *Usability (User) Testing 101*. Nielsen Norman Group. <https://www.nngroup.com/articles/usability-testing-101/>
- Sarkar, D. (DJ). (2018, September 13). *A Comprehensive Guide to the Grammar of Graphics for Effective Visualization of Multi-dimensional....* Medium. <https://towardsdatascience.com/a-comprehensive-guide-to-the-grammar-of-graphics-for-effective-visualization-of-multi-dimensional-1f92b4ed4149>
- Sherry Seethaler (Regisseur). (2017, September 14). *Now You See It: Lessons from Research on Perception for Design of Data Visualizations* [Video recording]. <https://www.youtube.com/watch?v=66eE4rc5xU0>
- Sibinga, E., Waldron, E., & Waldron, E. S. & E. (2021, September 30). Cognitive Load as a Guide: 12 Spectrums to Improve Your Data Visualizations, Nightingale. *Nightingale*. <https://nightingaledvs.com/cognitive-load-as-a-guide-12-spectrums-to-improve-your-data-visualizations/>
- Todorovic, D. (2008). Gestalt principles. *Scholarpedia*, 3(12), 5345. <https://doi.org/10.4249/scholarpedia.5345>

Vopani. (2023, November 7). *Formula 1 World Championship (1950—2024)*. Kaggle.

<https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020>

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, 4(1),

301–350. <https://doi.org/10.1007/BF00410640>

Wickham, H., & Grolemund, G. (2017). *R for Data Science*. <https://r4ds.had.co.nz/wrangle-intro.html>