

Modul GDV Abgabe



Bericht über die Grundlagen der Datenvisualisierung

Autor: Jan Zumwald

Fachexperte/-in: Arzu Cöltekin

Milena Rudig

Studiengang: B.Sc. Data Science

Modul: Grundlagen der Datenvisualisierung

Ort, Datum: Brugg-Windisch, 05.01.2025

Inhaltsverzeichnis

Abbildungsverzeichnis	3
1 Einleitung	4
2 LE 1 – Grundlagen der Visualisierung, Diagrammtypen	5
3 LE 2 – Visuelle Wahrnehmung	8
4 LE 3 – Designprinzipien vs. Daten	10
5 LE 4 – Grammatik der Grafik	13
6 LE 5 - Evaluation	15
7 Schlusswort	17
Literaturverzeichnis	18
Umfrage	19
Anhang	22

Abbildungsverzeichnis

Abbildung 1: Verteilung der Gesamttore pro Spiel in der Premier League	4
Abbildung 2: Tore und Gegentore von Arsenal pro Spieltag	6
Abbildung 3: Passgenauigkeit vs. Tordifferenz von Arsenal.....	7
Abbildung 4: Ballbesitz vs. Tordifferenz von Arsenal.....	7
Abbildung 5: Tore und Gegentore von Arsenal pro Spieltag (Negativbeispiel)	8
Abbildung 6: Zusammenhang zwischen Fouls, Gelben Karten und Punkten in der Premier League	9
Abbildung 7: Top 3 und schlechteste 3 Teams nach Ballbesitz.....	10
Abbildung 8: Top 3 und schlechteste 3 Teams nach Gesamtpunkten	10
Abbildung 9: Passgenauigkeit von Arsenal im Verlauf der Saison	11
Abbildung 10: Erfolgsrate Tacklings von Arsenal im Verlauf der Saison.....	11
Abbildung 11: Ballbesitz von Arsenal im Verlauf der Saison (2-dimensional).....	13
Abbildung 12: Ballbesitz von Arsenal im Verlauf der Saison (3-dimensional).....	13
Abbildung 13: Ballbesitz von Arsenal im Verlauf der Saison (4-dimensional).....	14
Abbildung 14: Ballbesitz von Arsenal im Verlauf der Saison (5-dimensional).....	14
Abbildung 15: Durchschnittliche Fouls, durchschnittliche gelbe Karten und Punkte pro Team tabellarisch dargestellt	27
Abbildung 16: Top 3 und schlechteste 3 Teams nach Tordifferenz.....	27
Abbildung 17: Top 3 und schlechteste 3 Teams nach Passgenauigkeit	28
Abbildung 18: Passgenauigkeit vs. Gesamtpunkte pro Team	28

1 Einleitung

Im Modul "Grundlagen der Datenvisualisierung" sollen über fünf Lernergebnisse die Prinzipien erlernt werden, um Daten effektiv und verständlich zu visualisieren. Dabei werden die verschiedenen Grafiktypen thematisiert sowie deren Wahrnehmung, aber auch die Schritte, die der eigentlichen Visualisierung vorausgehen, wie die Datenaufbereitung und deren Einfluss auf das Visualisierungsdesign. Ein zentrales Konzept ist zudem die "Grammatik der Grafik", welches dabei hilft, Visualisierungen systematisch aufzubauen. Abschliessend soll auch die effiziente Evaluation von Visualisierungen erlernt werden. (*Porträt - (gdv) Fundamentals of Data Visualization - Spaces, 2024*)

Für den Bericht über die Grundlagen der Datenvisualisierung wurde als Basis ein Datensatz gewählt, der alle EPL (English Premier League) Spiele der Saison 2022/2023 umfasst (*Premier League 2022-23, 2023*). Der Datensatz enthält eine Vielzahl von Variablen, von den geschossenen Toren und der Ballbesitzverteilung bis hin zur Anzahl der Zuschauer. Ziel ist es, anhand des Datensatzes sowohl die theoretischen Lerninhalte des Moduls anzuwenden als auch einen detaillierten Einblick in die Saison der Premier League zu gewinnen. Ein besonderer Fokus liegt darauf, mögliche Gründe dafür zu identifizieren, weshalb Arsenal trotz langer Tabellenführung die Meisterschaft am Ende nicht gewinnen konnte. Dabei sollen Visualisierungen erstellt werden, die Fussballinteressierten die wichtigsten Informationen verständlich präsentieren. Die Grafiken werden für unterschiedliche Anspruchsniveaus gestaltet und anschliessend auf ihre Wirksamkeit geprüft.

2 LE 1 – Grundlagen der Visualisierung, Diagrammtypen

Mit Visualisierungen können Daten einfach dargestellt werden, wobei die Wahl eines passenden Diagrammtyps entscheidend ist. Zu Beginn sollte man sich am besten Fragen, was der lesenden Person gezeigt werden soll. Dies kann zum Beispiel ein Vergleich von zwei oder mehreren Variablen, eine Abhängigkeit von zwei Variablen oder eine Verteilung von nur einer Variable sein. Sobald dies klar ist, muss zusätzlich berücksichtigt werden, wer das Publikum ist und welche Art von Diagramm für diese am einfachsten verständlich ist. (*Chart Chooser*, 2013)

Im Beispiel mit dem EPL-Datensatz soll einfach gestartet werden. Zu Beginn stellt sich die Frage, wie die Verteilung der Tore pro Spiel aussieht, also wie viele Spiele gab es ohne Tore, mit einem Tor, mit zwei Toren usw. gab. Da Sportstatistiken und -diagramme oft ein breites Publikum ansprechen, wird besonderen Wert daraufgelegt, die Visualisierungen möglichst verständlich und zugänglich zu gestalten. Das Ziel ist also, eine univariate Verteilung so einfach wie möglich darzustellen. Hierfür wird optimalerweise ein Histogramm verwendet.

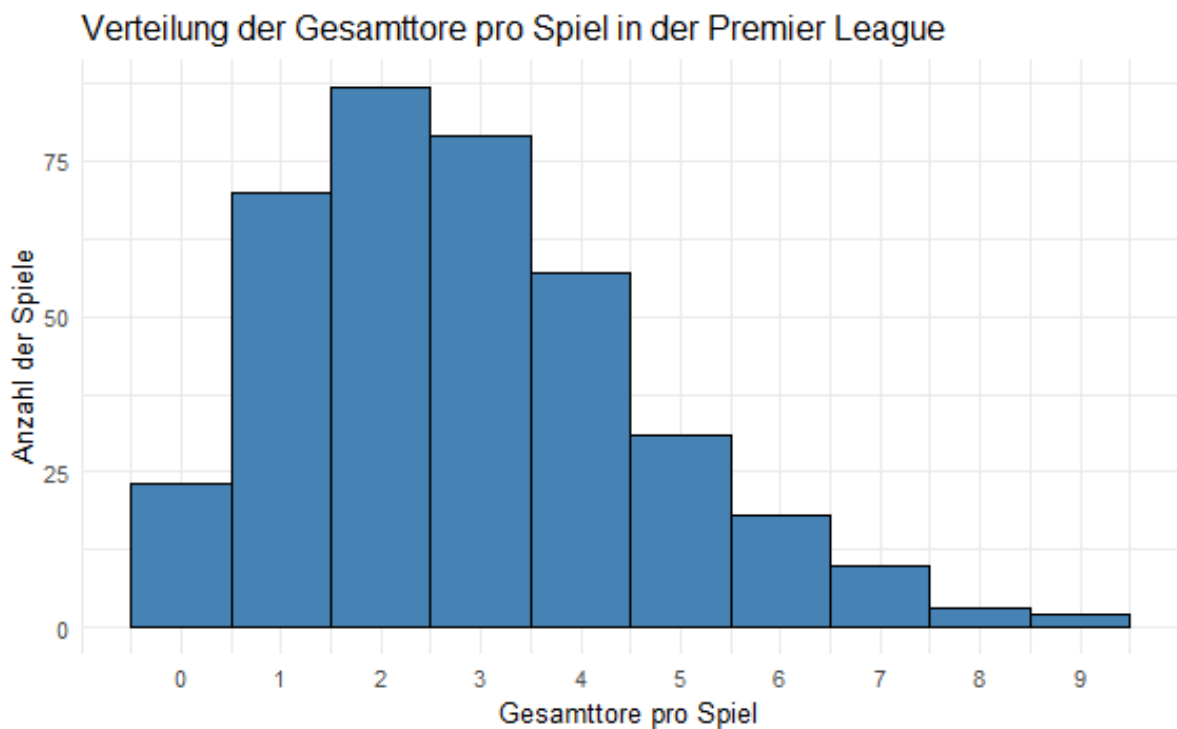


Abbildung 1: Verteilung der Gesamttore pro Spiel in der Premier League

In Abbildung 1 ist durch die hohen Balken direkt ersichtlich, dass in den meisten Spielen 1 bis 3 Tore gefallen sind aber mit der abfallenden Kurve auch, dass es einige wenige Spiele gab mit vielen Toren. Das zuvor erwähnte Ziel, die Verteilung der Anzahl Tore pro Spiel darzustellen, konnte mit dem Histogramm entsprechend erreicht werden. Dies soll anschliessend tiefergehend angeschaut werden, da mit dem vorderen Diagramm zwar bereits eine gute Übersicht gewonnen werden konnte, aber noch nicht mehr.

In einem zweiten Punkt sollen für ein bestimmtes Team die geschossenen und erhaltenen Tore pro Spiel visualisiert werden. So kann ein tiefer Einblick in die Performance dieses Teams durch die Saison gewonnen werden. In der Abbildung 2 ist das Ziel somit, zwei verschiedene Variablen (erzielte Tore und Gegentore) über die Zeit (Spieltag) zu visualisieren. Für diesen Anwendungsfall kann ein Liniendiagramm von grossem Nutzen sein.

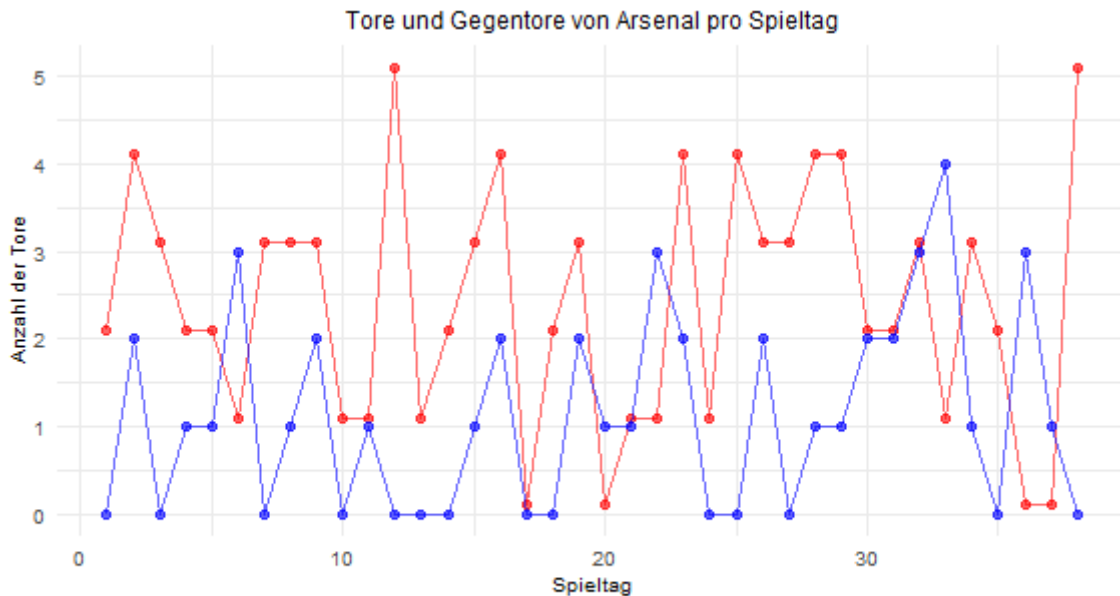


Abbildung 2: Tore und Gegentore von Arsenal pro Spieltag

Abbildung 2 beinhaltet alle Informationen, welche im zuvor festgelegten Ziel verlangt wurden. Da es den Verlauf zweier Variablen über 38 Spieltage darstellen soll, ist es bereits etwas komplexer als das Vordere. Bei längerem Auseinandersetzen mit der Abbildung 2 können viele Erkenntnisse gewonnen werden. Wenn die geschossenen Tore höher sind als die Gegentore – also der rote Datenpunkt deutlich über dem blauen Datenpunkt ist – bedeutet dies ein Sieg. Dem Diagramm kann also entnommen werden, dass Arsenal nach einer starken Startphase Mitte Saison eingestürzt ist und zwei unentschieden und zwei Niederlagen einfahren mussten. Sie konnten sich danach wieder fassen, schlossen die Saison jedoch wiederum nur mässig gut ab. Der Distanz zwischen den Punkten kann ausserdem entnommen werden, wie hoch die Tordifferenz im Spiel war. Bei den Spieltagen 12, 25 und 38 scheint es sich beispielsweise um klare Siege seitens Arsenals zu handeln.

In einem letzten Punkt sollen noch Abhängigkeiten untersucht werden. Die Frage hierbei ist, ob möglicherweise der Ballbesitz oder die Passgenauigkeit in dieser Saison entscheidend waren für ein erfolgreiches Spiel. Um Beziehungen zwischen zwei Variablen darzustellen, eignen sich Scatterplots hervorragend.

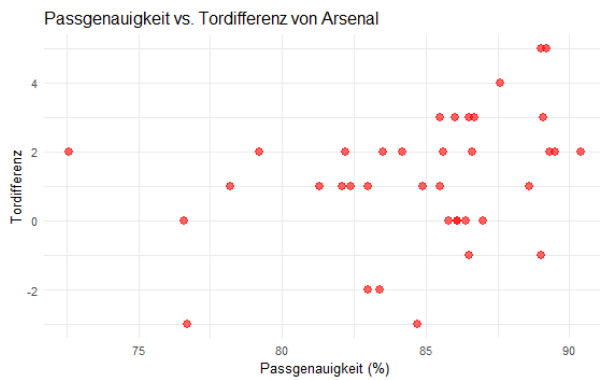


Abbildung 3: Passgenauigkeit vs. Tordifferenz von Arsenal

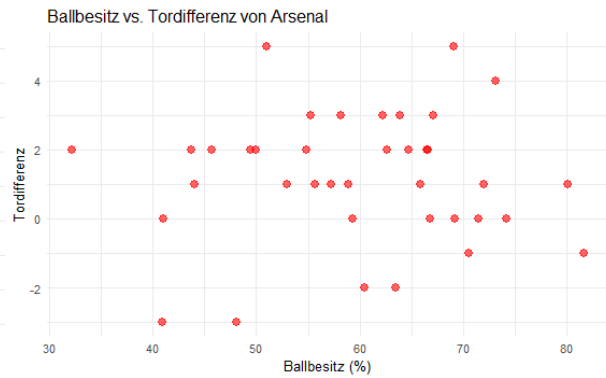


Abbildung 4: Ballbesitz vs. Tordifferenz von Arsenal

In den Abbildungen 3 und 4 steht jeweils ein Datenpunkt für ein Spiel. Hierbei wird der Zeitverlauf nicht inkludiert, dies soll in diesem Fall aber auch nicht im Vordergrund stehen. Beim Ballbesitz in der Abbildung 4 sind weniger klare Muster ersichtlich als bei der Passgenauigkeit in der Abbildung 3. Es wurden einige Spiele gewonnen, bei denen Arsenal weniger Ballbesitz als der Gegner hatte, es gab aber auch Niederlagen, bei denen Arsenal ein Ballbesitz von über 60% hatte. Auch im Scatterplot in der Abbildung 3 mit der Passgenauigkeit gibt es diverse Ausreisser, man kann aber erkennen, dass hohe Passgenauigkeiten in dieser Saison möglicherweise zu höheren Tordifferenzen beigetragen hatten. Ein Fokus darauf könnte also die Performance des ganzen Teams fördern. Die Stärke dieser Korrelationen könnten ausserdem zusätzlich mit Berechnungen aufgezeigt werden, da hier der Fokus auf den Visualisierungen liegt wird darauf verzichtet.

3 LE 2 – Visuelle Wahrnehmung

Im zweiten Lernergebnis geht es um die menschliche Wahrnehmung von Visualisierungen. Es wird unterstrichen, wie wichtig es ist, ein Verständnis dafür zu entwickeln, um Graphen und Designs so informativ und verständlich wie möglich zu gestalten. Eine zentrale Thematik der visuellen Wahrnehmung sind die sechs Prinzipien der Gestalttheorie. Sie sollen dabei dienen, Graphen und Diagramme so darzustellen, dass sie das visuelle Verarbeitungsnetzwerk des Menschen ansprechen. (Schwabisch, 2021)

Dabei handelt es sich um die folgenden Prinzipien:

- Nähe: Objekte in räumlicher Nähe werden als Gruppe wahrgenommen.
- Ähnlichkeit: Ähnliche Objekte (in Farbe, Form) bilden eine Gruppe.
- Einschluss: Umschlossene Elemente werden als zusammengehörig erfasst.
- Schliessung: Unvollständige Formen werden vom Gehirn vervollständigt.
- Kontinuität: Alignierte oder fortgesetzte Objekte gelten als Einheit.
- Verbindung: Verbundene Elemente werden als Teil derselben Gruppe gesehen.

Bei dem Punkt der Ähnlichkeit können ausserdem noch die sieben visuellen Variablen nach Jaques Bertin berücksichtigt werden. Auch diese Variablen sollen zusätzlich als Bausteine dienen, um Visualisierungen verständlich aufzubauen. Dabei handelt es sich um: Position, Grösse, Form, Wert (Helligkeit), Farbe, Orientierung/Richtung und Textur. (*Visual Variables*, o. J.)

Als Beispiel soll die Abbildung 2 aus LE 1 dienen, wo die geschossenen Tore mit den Gegentoren von Arsenal gegenübergestellt wird. Hier werden die beiden Gruppen mit Verbindung (Liniendiagramm) und Ähnlichkeit (Einfärbung) dargestellt, was für das Publikum schnell klar sein sollte. Werden diese beiden Variablen entfernt, bleibt wie in Abbildung 5 ersichtlich nur noch ein einfacher Scatterplot, aus welchem nur wenig herausgelesen werden kann.

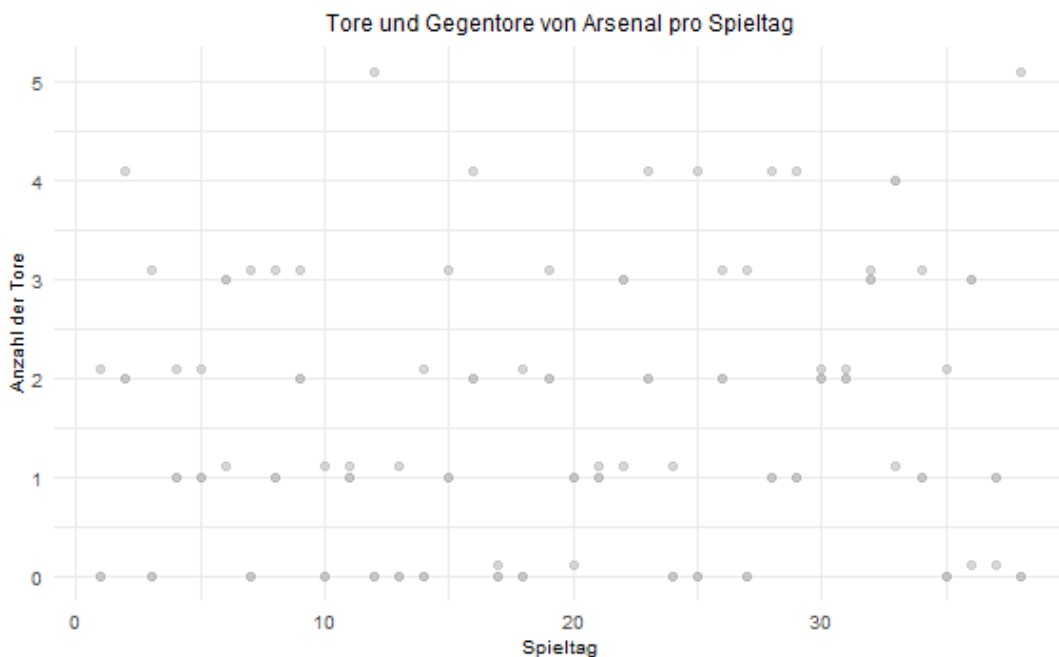


Abbildung 3: Tore und Gegentore von Arsenal pro Spieltag (Negativbeispiel)

Die Punkte in der Abbildung 5 werden nun im Vergleich zur Abbildung in Graustufen dargestellt und die Legende dazu wird nicht visualisiert. Dies macht es für die Lesenden herausfordernd, geschossene Tore und erhaltene Tore zu unterscheiden. Zudem fehlt die Verbindungslinie, welche zum einen unterstützt, den Ablauf der Saison darzustellen und zum anderen die Unterschiede von Spiel zu Spiel verdeutlicht.

In einem weiteren Schritt soll untersucht werden, ob ein Zusammenhang zwischen der Aggressivität der Mannschaften und den erzielten Punkten besteht. Als Indikatoren für die Aggressivität dient die durchschnittliche Anzahl an Fouls und gelben Karten. Da der Datensatz überschaubar ist, lässt sich eine erste Analyse bereits durch eine Tabelle durchführen (s. Anhang, Abbildung 15), um mögliche Muster zu erkennen. Doch eine Grafik bietet einen wesentlich besseren Überblick und erlaubt es, Muster und Zusammenhänge klarer hervorzuheben.

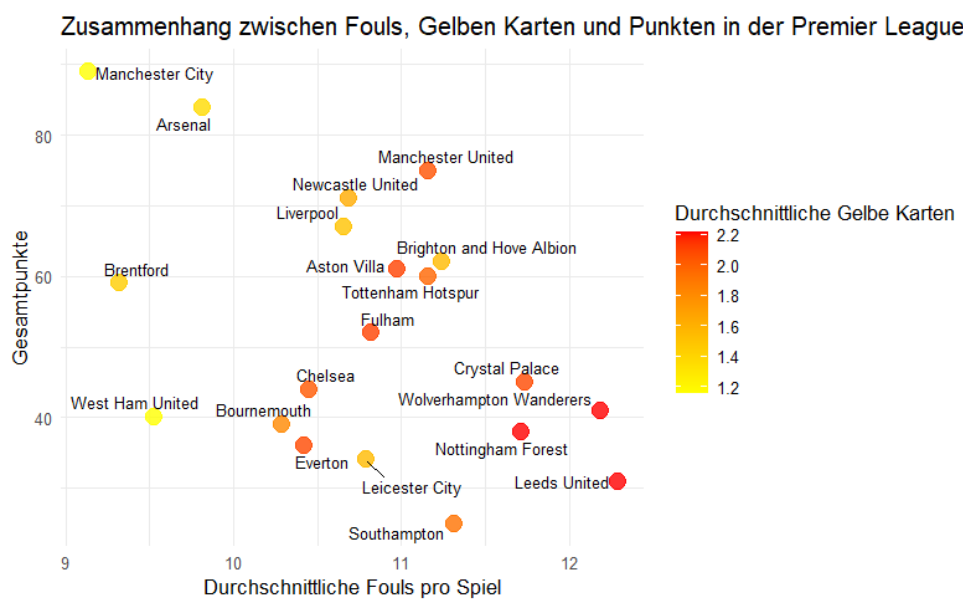


Abbildung 4: Zusammenhang zwischen Fouls, Gelben Karten und Punkten in der Premier League

In der Abbildung 6 wird erkennbar, dass Mannschaften, die sich horizontal nahe sind, eine ähnliche Anzahl an Fouls aufweisen, was auf das Gestaltprinzip der Nähe verweist. Dies erleichtert es dem Betrachter, Teams mit ähnlichem aggressiven Spielstil sofort zu gruppieren. Teams, die vertikal nah beieinander liegen, deuten auf Tabellennachbarn hin, was den Punktestand visualisiert und einen direkten Vergleich ermöglicht. Durch die farbliche Codierung, die die durchschnittliche Anzahl der Gelben Karten darstellt, wird das Gestaltprinzip der Ähnlichkeit genutzt: Teams mit einer ähnlichen Anzahl an Gelben Karten erscheinen in ähnlichen Farbtönen. Dies ermöglicht es, Muster zu erkennen, wie aggressivere Teams (mit vielen Gelben Karten) im Vergleich zu weniger aggressiven Teams abschneiden. Dabei wechseln die Punkte von Gelb in Rot, wenn Teams verhältnismässig durchschnittlich viele Gelbe Karten erhalten haben.

Diese Anwendung der visuellen Variablen – Position, Farbe und Nähe – folgt den Prinzipien der Gestalttheorie und sorgt dafür, dass komplexe Daten intuitiv und effizient wahrgenommen werden können. Durch die gezielte Anwendung dieser Prinzipien wird die Grafik zu einem leistungsstarken Werkzeug, um verborgene Muster zu entdecken und die Zusammenhänge zwischen aggressivem Spiel und Punkten zu analysieren.

4 LE 3 – Designprinzipien vs. Daten

In diesem Lernziel geht es darum, die Verbindung zwischen der Datenaufbereitung und den Designentscheidungen für Visualisierungen herzustellen. Oftmals liegen Daten in unstrukturierter Form vor und müssen zunächst in eine geeignete Struktur gebracht werden, um sie visualisieren zu können. Dabei müssen wichtige Entscheidungen getroffen werden, wie zum Beispiel die Klassifizierung von kategorialen Daten, der Umgang mit Unsicherheiten und die Behandlung von fehlenden Werten. Diese Entscheidungen beeinflussen das Design der Visualisierungen deutlich. Es ist entscheidend, dass die gewählten Designprinzipien den Eigenschaften der Daten entsprechen. (Elgabry, 2019)

Im Beispiel der Premier League liegen die Daten zu Beginn so vor, dass jedes Spiel ein einzelner Datenpunkt ist, wobei die Heim- und Auswärtsteams mit ihren Werten aufgeführt werden. Um jedoch eine allgemeine Übersicht über die Saison zu erhalten, bedarf es einer Aggregation der Daten. Daher wurde eine Zusammenfassung erstellt, die grundsätzlich die Ligatabelle am Ende der Saison repräsentiert, aber zusätzlich den Mittelwert des Ballbesitzes sowie der Passgenauigkeit über alle Spiele pro Team berücksichtigt. Aus dem ursprünglichen Datensatz, in dem jedes Spiel ein Datenpunkt ist, entsteht so ein neuer Datensatz, in dem jedes Team ein Datenpunkt ist und die Variablen die Kennzahlen der gesamten Saison darstellen. Mit Hilfe dieser Aggregation können bereits erste Einblicke in die Saison gewonnen werden, wie beispielsweise die Ermittlung der besten und schlechtesten Teams nach Punkten, Tordifferenz, durchschnittlichem Ballbesitz und durchschnittlicher Passgenauigkeit (siehe Anhang, zusätzliche Abbildungen 16 und 17).

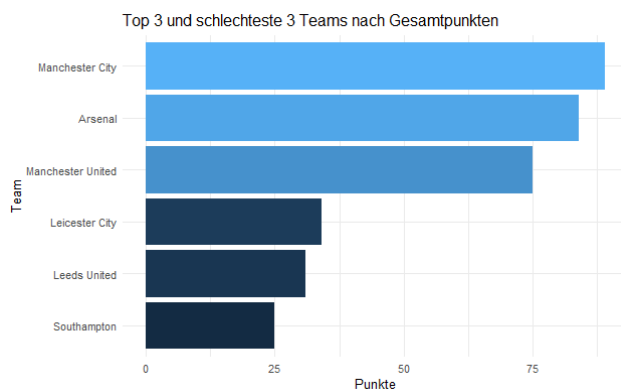


Abbildung 7: Top 3 und schlechteste 3 Teams nach Gesamtpunkten

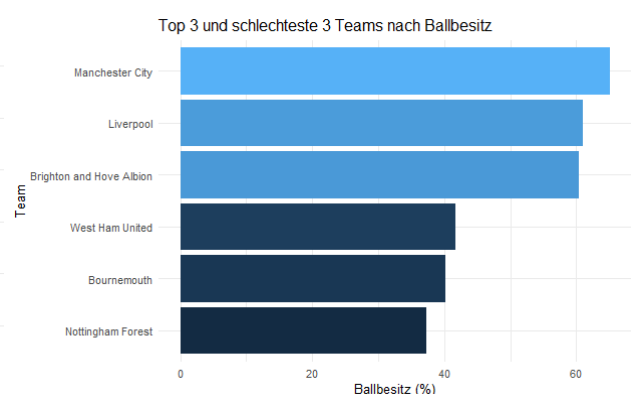


Abbildung 8: Top 3 und schlechteste 3 Teams nach Ballbesitz

Aus den Abbildungen 7 und 8 sowie 16 und 17 wird ersichtlich, dass Manchester City in allen Faktoren führend ist, und so wird die dominante Saison verstärkt beleuchtet. Zeitgleich ist interessant zu sehen, dass Nottingham Forest die dritt tiefste Tordifferenz hat und ausserdem am schlechtesten abschliesst in der Passgenauigkeit sowie beim Ballbesitz. Dennoch waren sie punktemässig nicht unter den schlechtesten 3 Teams und konnten so den Ligarhalt mit möglicherweise etwas Glück sichern. Wenn die Teams geplottet werden nach Passgenauigkeit und Gesamtpunkten wird zwar ein leichter linearer Trend ersichtlich (siehe Anhang, Abbildung 18), durch die Streuung wird aber klar, dass keine direkte Schlussfolgerung gezogen werden kann. Im Vergleich der Top 3 und schlechtesten 3 Teams wird diese Erkenntnis bestärkt, da nur wenige Teams in allen Grafiken übereinstimmen.

In einem weiteren Punkt sollte die Leistung von Arsenal durch die Saison evaluiert werden und weshalb es im Ende trotz langer Tabellenführung nicht ausgereicht hat für den Meistertitel. Da im Datensatz die Teams jeweils als Heim- oder Auswärtsteam mit den jeweiligen Leistungsdaten aufgeführt sind, können die Spiele eines spezifischen Teams nicht direkt analysiert werden. Die Daten müssten umgeformt werden, so dass jedes Spiel in der Sicht von Arsenal vorhanden ist inkl. den Daten des Gegners. Mithilfe der Umformung kann nun der Saisonverlauf zum Beispiel anhand Ballbesitz oder Anzahl Fouls pro Spiel dargestellt werden.

In den beiden Grafiken ist der Saisonverlauf nach Passgenauigkeit und nach Tackling Erfolgsrate dargestellt. Ausserdem sind die Spiele farblich gekennzeichnet nach Sieg, Niederlage oder Unentschieden.

In der Passgenauigkeit wird anhand der y-Achse ersichtlich, dass Spiele mit einer sehr guten Passgenauigkeit (um die 90%) beinahe immer gewonnen wurden. In den Spielen mit durchschnittlicheren Passgenauigkeiten wird kein Muster ersichtlich. Es wurde auch ein Spiel gewonnen, wo die Passgenauigkeit 75% war, also jeder vierte Pass ging in diesem Spiel zum Gegner.

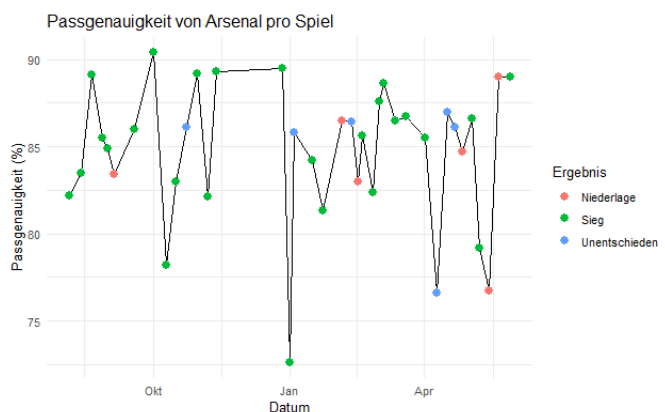


Abbildung 5: Passgenauigkeit von Arsenal im Verlauf der Saison

Auch in der Abbildung 10 gibt es interessante Einblicke. Es gab nur 5 Spiele mit einer Tackling Erfolgsrate von 40% oder weniger. Von diesen 5 Spielen wurden 4 verloren. Zudem gab es in der Saison nur 6 Niederlagen in 38 Spielen. Mit dem Trainieren von konstanterer Defensivarbeit könnten folglich möglicherweise Niederlagen vermieden werden.

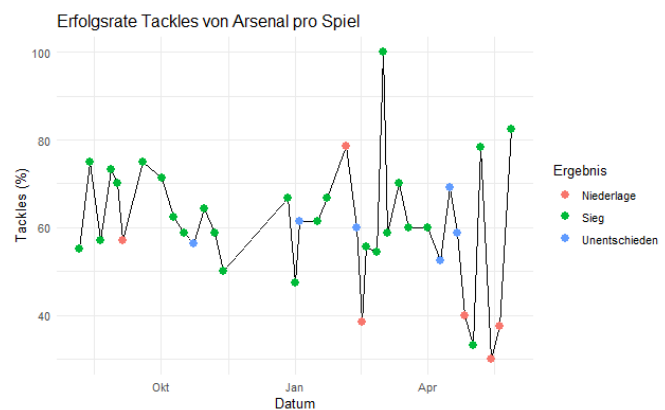


Abbildung 6: Erfolgsrate Tacklings von Arsenal im Verlauf der Saison

Die Analyse der Premier-League-Saison zeigt deutlich die enge Verbindung zwischen der Datenaufbereitung und den Designprinzipien. Um Arsenals Spiele sinnvoll analysieren zu können, mussten die Spieldaten so umstrukturiert werden, dass Muster und Trends, wie der Zusammenhang zwischen Passgenauigkeit, Tackling Erfolgsrate und Spielergebnissen, erkennbar wurden.

Datenentscheidungen wie die Aggregation auf Teamebene und die Umwandlung einzelner Spiele in Metriken hatten direkten Einfluss darauf, welche Einblicke gewonnen und wie die Informationen visualisiert werden konnten. Ohne diese Umstrukturierung wären die Erkenntnisse nur schwer zugänglich gewesen. Die Anwendung grundlegender Designprinzipien wie Klarheit und Vergleichbarkeit ermöglichte es, die relevanten Metriken über die gesamte Saison hinweg zu analysieren. Damit wird deutlich, dass die Qualität einer Visualisierung stark von der richtigen Datenaufbereitung abhängt und dass durchdachte Designentscheidungen die Komplexität der Daten verständlich machen.

5 LE 4 – Grammatik der Grafik

Die Grammatik der Grafik (Grammar of Graphics) ist ein strukturierter Ansatz zur Erstellung von Datenvisualisierungen, der es ermöglicht, mehrdimensionale Daten effektiv und übersichtlich darzustellen. Ursprünglich von Leland Wilkinson entwickelt und später von Hadley Wickham in der ggplot2-Bibliothek popularisiert (Wickham et al., 2023), ermöglicht dieser Ansatz eine klare Trennung der verschiedenen Elemente einer Grafik, wie Daten, Ästhetik, geometrische Objekte und Skalen. Dies hilft dabei, komplexe Datensätze besser zu verstehen, indem verschiedene Dimensionen der Daten gleichzeitig sichtbar gemacht werden. Durch die Verwendung von Farbcodierungen, Grössenunterschieden und Facetten können mehrdimensionale Daten in einer zweidimensionalen Darstellung sinnvoll visualisiert werden. Die Grammatik der Grafik bietet somit einen flexiblen und erweiterbaren Rahmen, um aussagekräftige und anpassungsfähige Visualisierungen zu erstellen. (Sarkar, 2018)

Durch die Verwendung von Farbcodierungen, Grössenunterschieden und Facetten können mehrdimensionale Daten in einer zweidimensionalen Darstellung verständlich visualisiert werden. Die Grammatik der Grafik bietet somit nicht nur einen flexiblen und erweiterbaren Rahmen, um aussagekräftige Visualisierungen zu erstellen, sondern auch ein Werkzeug, um tiefere Einblicke in die Daten zu gewinnen, die durch rein statistische Auswertungen oft verborgen bleiben. Anhand der Grammatik der Grafik soll nun Schritt für Schritt beobachtet werden, wie der Saisonverlauf von Arsenal in einer zweidimensionalen Grafik mit mehrdimensionalen Daten sinnvoll dargestellt werden kann.

Daten: Die Visualisierung beginnt mit den grundlegenden Daten – in der Abbildung 11 mit dem Datum und dem Ballbesitz. Diese beiden Variablen bilden die Grundlage der Visualisierung und definieren, welche Informationen dargestellt werden sollen. Auf der X-Achse steht das Datum der Spiele, auf der Y-Achse der Ballbesitz in Prozent.

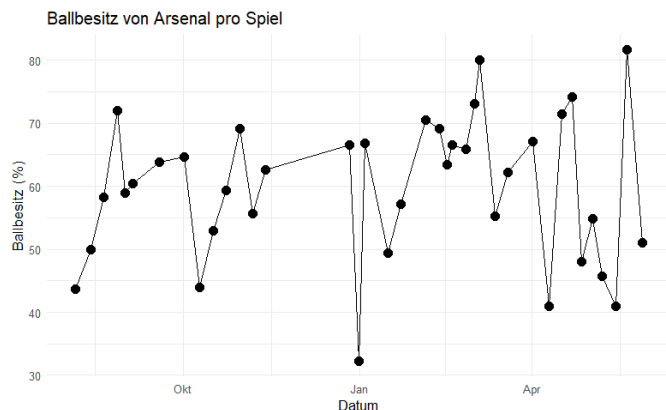


Abbildung 7: Ballbesitz von Arsenal im Verlauf der Saison (2-dimensional)

Ästhetik: Im zweiten Schritt wird eine dritte Dimension hinzugefügt, das Ergebnis des Spiels (Sieg, Unentschieden, Niederlage). Die Ästhetik wird hier verwendet, um das Spielergebnis mithilfe von Farben darzustellen. Diese Farbkodierung macht es dem Betrachter leicht, die Resultate der Spiele auf einen Blick zu erfassen.

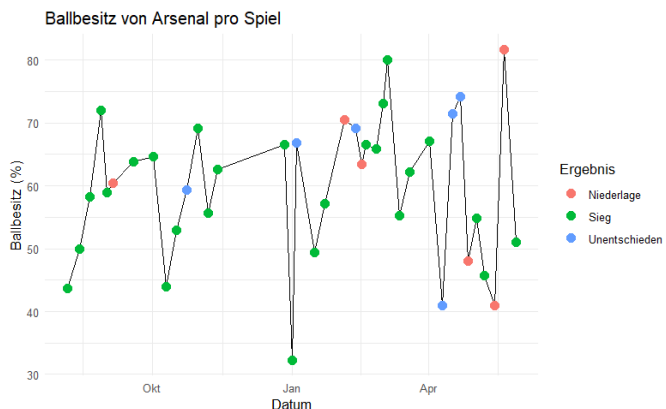


Abbildung 10: Ballbesitz von Arsenal im Verlauf der Saison (3-dimensional)

Skalierung: Im nächsten Schritt kommt die Tackling-Erfolgsrate hinzu. Diese wird durch die Grösse der Punkte dargestellt. Hier kommt das Konzept der Skalierung ins Spiel: Die Grösse der Punkte skaliert proportional zur Tackling Erfolgsrate. Auf diese Weise kann eine vierte Dimension in die Grafik integriert werden, ohne die visuelle Klarheit zu beeinträchtigen.

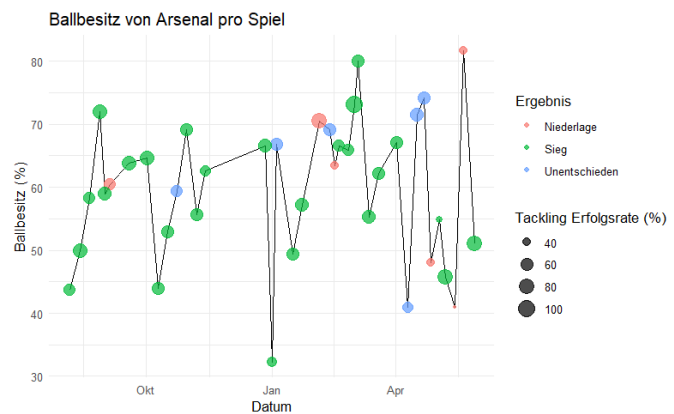


Abbildung 11: Ballbesitz von Arsenal im Verlauf der Saison (4-dimensional)

Geometrische Objekte: Die Visualisierung verwendet unterschiedliche geometrische Objekte, um den Spieltyp zu unterscheiden. Für Heimspiele werden Dreiecke verwendet, während für Auswärtsspiele Punkte genutzt werden. Diese Formkodierung hilft, den Kontext der Spiele besser zu verstehen, indem sofort sichtbar ist, ob es sich um ein Heim- oder Auswärtsspiel handelt.

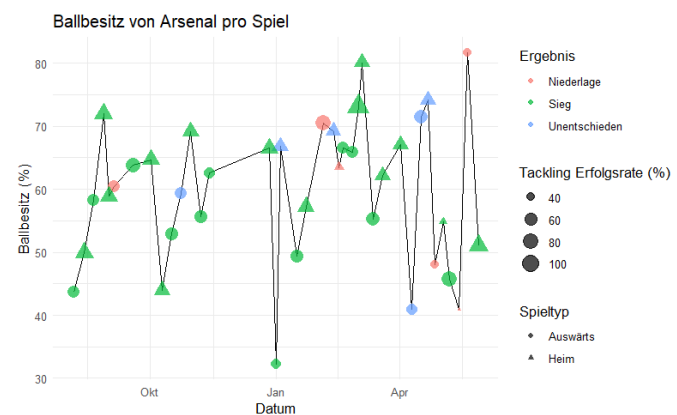


Abbildung 14: Ballbesitz von Arsenal im Verlauf der Saison (5-dimensional)

Durch diesen schrittweisen Aufbau der Visualisierung wird der Ansatz der Grammatik der Grafik deutlich: Es wird eine klare Struktur geschaffen, bei der jede Ebene (Data, Aesthetics, Scale, Geometric Objects, etc.) methodisch hinzugefügt wird, um die mehrdimensionalen Daten in einem zweidimensionalen Raum darzustellen. Im letzten Schritt wirkt die Abbildung 14 überladen und läuft der Gefahr entgegen, eine Kernaussage zu verlieren. Es wird klar, dass Entscheidungen getroffen werden müssen und dass weniger manchmal mehr ist.

Zusätzlich könnten noch Facetten hinzugefügt werden, um z. B. den Saisonverlauf in Hinrunde und Rückrunde aufzuteilen, oder statistische Elemente wie Durchschnittslinien und Trends integriert werden, um tiefergehende Muster in den Daten zu erkennen. Da die Grafik jedoch bereits eine Vielzahl an unterschiedlichen Informationen wie Spielergebnis, Tackling Erfolgsrate und Spieltyp enthält, wird bewusst darauf verzichtet, um die Übersichtlichkeit zu bewahren. Auf diese Weise bleibt die Visualisierung trotz der Mehrdimensionalität klar und leicht verständlich.

6 LE 5 - Evaluation

Die Evaluation von Visualisierungen ist ein entscheidender Schritt im Designprozess, um sicherzustellen, dass die präsentierten Daten nicht nur korrekt, sondern auch verständlich für die Zielgruppe sind. Visualisierungen dienen dazu, komplexe Informationen in leicht zugängliche Grafiken zu übersetzen. Doch selbst die besten Daten können in einer schlecht gestalteten Visualisierung ihre Wirkung verlieren, wenn sie für die Nutzer schwer zu interpretieren oder missverständlich sind. Durch eine gezielte Analyse der Nutzererfahrung und -wahrnehmung können Visualisierungen optimiert werden, um sowohl inhaltlich als auch visuell effektiv zu sein.

Christian Rohrer (Rohrer, 2008) beschreibt die verschiedenen Methoden zur Erforschung des Nutzererlebnisses (User Experience, UX) und erläutert, wann welche Methode verwendet werden sollte. Die Methoden lassen sich in drei Hauptdimensionen unterscheiden:

Einstellungsbezogen vs. Verhaltensbezogen:

- **Einstellungsbezogen:** Hier geht es um das, was die Nutzer über ein Produkt sagen (z. B. in Umfragen oder Fokusgruppen), um ihre Meinungen und Vorlieben zu erfassen. Diese Methoden sind nützlich, um Ansichten zu messen oder zu verstehen.
- **Verhaltensbezogen:** Hier wird untersucht, was die Nutzer tatsächlich tun. Diese Methoden, wie Usability-Studien oder Eyetracking, konzentrieren sich auf das tatsächliche Verhalten der Nutzer, um Probleme in der Nutzung zu identifizieren.

Qualitativ vs. Quantitativ:

- **Qualitative Methoden:** Diese zielen darauf ab, das "Warum" hinter dem Nutzerverhalten zu verstehen. Methoden wie Feldstudien oder Usability-Tests liefern tiefere Einblicke in die Art und Weise, wie Nutzer ein Produkt verwenden, und decken verborgene Probleme auf.
- **Quantitative Methoden:** Diese messen das "Wie viel" oder "Wie oft". Sie liefern numerische Daten und Metriken, wie etwa in A/B-Tests oder Benchmarking-Studien, um die Effektivität von Designänderungen zu bewerten.

Kontext der Produktnutzung:

- **Natürliche Nutzung:** Hier wird die Produktnutzung unter realen Bedingungen untersucht, z. B. durch ethnografische Feldstudien.
- **Vorgegebene Nutzung:** In Laboren durchgeführte Tests, wie Usability-Studien, legen eine bestimmte Nutzung vor, um gezielt bestimmte Interaktionen zu analysieren.
- **Keine Produktnutzung:** Methoden wie Markenstudien oder Befragungen, bei denen das Produkt selbst nicht verwendet wird.

Für die Evaluation der Grafiken wird Usability Testing verwendet. Das Usability Testing kombiniert vor allem qualitative und verhaltensbezogene Methoden, da es darauf abzielt, das tatsächliche natürliche Nutzerverhalten zu beobachten und zu analysieren. Es ermöglicht Einblicke in das „Warum“ hinter den Schwierigkeiten der Nutzer:innen und deckt mögliche Missverständnisse oder Barrieren in der Nutzung der Visualisierungen auf. Diese Methode ist ideal, um herauszufinden, ob die Visualisierungen verständlich sind und den Nutzern die gesuchten Informationen vermitteln. Usability Testing ermöglicht es, Probleme schnell zu identifizieren und Verbesserungspotenzial zu erkennen.

Durch das Testen mit 5 Teilnehmenden lassen sich oftmals die meisten Usability-Probleme bereits aufdecken, was diese Methode auch effektiv macht. (Moran, 2019)

Usability Testing - Umfrage

Die Teilnehmenden der Umfrage sollen sich mit den drei Abbildungen 1, 6 und 14 auseinandersetzen und folgende Fragen beantworten:

- Was fällt dir auf den ersten Blick auf?
- Wie würdest du das Hauptthema oder die Hauptaussage dieser Visualisierung beschreiben?
- Wie leicht fällt es dir, die Informationen aus dieser Visualisierung zu vergleichen und zu interpretieren? Gibt es Bereiche, die verwirrend oder unklar sind?

Da die Visualisierungen als Zielpublikum Fussballinteressenten hat, wurden in der Umfrage 5 Hobbyfussballer befragt, welche wenige bis durchschnittliche statistische Kenntnisse besitzen. Nachfolgend werden die Rückmeldungen der Umfrage zusammengefasst und dabei die Schlüsselerkenntnisse aufgeführt. Die detaillierte Umfrage wird dem Bericht beigelegt.

Test 1 (Abbildung 1): Verteilung der Tore pro Spiel

Die Visualisierung wurde von den meisten Nutzern als klar und leicht verständlich wahrgenommen. Der Höchstwert bei ein bis drei Toren pro Spiel fiel sofort auf, und die Struktur des Balkendiagramms wurde als intuitiv beschrieben. Die Hauptaussage – die häufigste Anzahl an Toren pro Spiel – war schnell erkennbar. Verbesserungen wurden kaum vorgeschlagen, da die Darstellung bereits als sehr übersichtlich empfunden wurde. Die Einfachheit des Designs machte es den Nutzern leicht, die Daten schnell zu erfassen und zu vergleichen.

Test 2 (Abbildung 6): Zusammenhang zwischen Fouls, Gelben Karten und Punkten

Die Darstellung des Zusammenhangs zwischen Fouls, Gelben Karten und Punkten wurde positiver bewertet, allerdings gab es einige Kritikpunkte. Die Farbkodierung für Gelbe Karten war für manche Nutzer schwer zu unterscheiden, besonders bei dicht beieinander liegenden Teams. Die Grafik vermittelt jedoch gut die Beziehung zwischen aggressivem Spiel und Punktzahl. Eine stärkere Differenzierung der Farben könnte die Übersicht verbessern, wodurch die Nutzer den Zusammenhang zwischen Fouls und Punkten noch klarer erkennen könnten.

Test 3 (Abbildung 14): Ballbesitz von Arsenal pro Spiel

Die Visualisierung des Ballbesitzes von Arsenal bot viele nützliche Informationen, wurde aber als komplex wahrgenommen. Die Symbole für Heim- und Auswärtsspiele sowie die Farbmarkierungen für Spielergebnisse wurden als hilfreich empfunden, während die Tackling Erfolgsrate, die durch Punktgrößen dargestellt wurde, nicht sofort ins Auge fiel. Eine Reduktion der dargestellten Dimensionen oder eine deutliche Hervorhebung der Tackling Erfolgsrate könnte die Lesbarkeit verbessern. Einige Nutzer fanden die Grafik etwas überladen, was den Vergleich der Informationen erschwerte, obwohl das Design insgesamt dennoch als hilfreich angesehen wurde.

Fazit des Usability Testings

Das durchgeführte Usability Testing hat bereits wertvolle Einblicke in die Stärken und Schwächen der Visualisierungen geliefert. Die Teilnehmer konnten wichtige Informationen aus den Grafiken extrahieren, und erste Verbesserungsvorschläge wurden sichtbar, besonders in Bezug auf die Klarheit der Farb- und Grössenkodierungen sowie die Lesbarkeit komplexerer Darstellungen. Dennoch war die Umfrage in ihrem Umfang begrenzt – sowohl was die Anzahl der gestellten Fragen als auch die der Teilnehmenden betrifft. Mit mehr Ressourcen hätte eine umfassendere Befragung durchgeführt werden können, die nicht nur mehr Teilnehmende, sondern auch detailliertere Fragen zur Interaktivität und Vergleichbarkeit der Daten umfassen könnte.

7 Schlusswort

Im abschliessenden Fazit zeigt sich, dass Arsenal trotz einer klaren Dominanz in der ersten Hälfte der Saison durch zwei schwächere Phasen die Tabellenführung abgeben musste. Diese Einbrüche in ihrer Leistung, vor allem in der Mitte und gegen Ende der Saison, waren entscheidend dafür, dass sie den Meistertitel nicht erringen konnten. Ein bemerkenswerter Faktor war dabei die Tackling-Erfolgsrate: Spiele, in denen Arsenal eine konstante und hohe Tackling-Rate aufwies, führten seltener zu Niederlagen. Eine Verbesserung in dieser defensiven Disziplin könnte in Zukunft dazu beitragen, Schwächen in kritischen Momenten der Saison zu vermeiden.

Manchester City hingegen gewann die Premier League durchgängig und verdient. Sie dominierten nicht nur in den entscheidenden Spielen, sondern auch in nahezu allen relevanten Statistiken wie Ballbesitz und Passgenauigkeit. Diese konstant hohe Leistung spiegelt sich sowohl in ihren Punkten als auch in ihrer Gesamtpformance wider. Damit wird klar, dass neben der individuellen Qualität auch die Stabilität über die gesamte Saison hinweg entscheidend für den Erfolg ist.

Literaturverzeichnis

Chart chooser. (2013, April 6). Storytelling with Data.

<https://www.storytellingwithdata.com/blog/2013/04/chart-chooser>

Elgabry, O. (2019, März 2). *The Ultimate Guide to Data Cleaning*. Medium.

<https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4>

Moran, K. (2019, Dezember 1). *Usability Testing 101*. Nielsen Norman Group.

<https://www.nngroup.com/articles/usability-testing-101/>

Porträt—(Gdv) Fundamentals of Data Visualization—Spaces. (2024, September 17).

<https://spaces.technik.fhnw.ch/spaces/fundamentals-of-data-visualization>

Premier League 2022-23. (2023, September 11). Kaggle.

<https://www.kaggle.com/datasets/afnanurrahim/premier-league-2022-23>

Rohrer, C. (2008, Oktober 6). *Wann bei der Erforschung von Nutzererlebnissen welche Methoden benutzt werden sollten*. usability.ch. <https://www.usability.ch/news.html?print=1&L=0>

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>

Schwabisch, J. (2021). *Better Data Visualizations*.

https://www.dropbox.com/scl/fi/b0a1b5i9p32zgy5cucm7d/schwabish_2021_chapter01.pdf?dl=0&rlkey=e1k389zz0b68d4djfu8dihrkp

Visual Variables. (o. J.). Abgerufen 18. März 2024, von <https://www.axismaps.com//guide/visual-variables>

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). 1 Data visualization. In *R for Data Science (2e)*. <https://r4ds.hadley.nz/>

Umfrage

Test 1:

Frage 1: Was fällt dir auf den ersten Blick auf?

1. Die Verteilung ist fast symmetrisch, mit einem klaren Höchstwert bei ein bis drei Toren pro Spiel.
2. Es gibt einen klaren Höhepunkt bei zwei Toren pro Spiel, und es ist offensichtlich, dass es nur sehr wenige Spiele mit null oder neun Toren gab.
3. Spiele mit zwei oder drei Toren sind am häufigsten. Spiele mit mehr als fünf Toren sind selten.
4. Der Bereich mit zwei bis drei Toren pro Spiel ist am auffälligsten, da er die meisten Spiele umfasst.
5. Die meisten Spiele haben zwei Tore, und es gibt einen deutlichen Abfall bei den Spielen mit mehr oder weniger Toren.

Frage 2: Wie würdest du das Hauptthema oder die Hauptaussage dieser Visualisierung beschreiben?

1. Es zeigt die Verteilung der Gesamttore pro Spiel in der Premier League. Die meisten Spiele enden mit ein bis drei Toren.
2. Die Grafik zeigt, wie viele Tore pro Spiel in der Premier League durchschnittlich erzielt werden, meist ein bis drei Tore.
3. Die Visualisierung zeigt, wie viele Tore in einem durchschnittlichen Premier-League-Spiel fallen.
4. Die Grafik zeigt, dass die meisten Spiele in der Premier League mit zwei oder drei Toren enden.
5. Die Verteilung der Tore in Premier-League-Spielen wird dargestellt, wobei zwei Tore am häufigsten sind.

Frage 3: Wie leicht fällt es dir, die Informationen aus dieser Visualisierung zu vergleichen und zu interpretieren? Gibt es Bereiche, die verwirrend oder unklar sind?

1. Sehr einfach und klar. Die Balkenhöhe macht es leicht, die Verteilung der Tore zu erkennen.
2. Leicht verständlich. Die Achsen und Daten sind gut beschriftet und leicht zu interpretieren.
3. Die Grafik ist sehr einfach zu interpretieren. Die Botschaft ist klar und prägnant.
4. Es ist leicht, die Informationen zu vergleichen. Die Darstellung ist intuitiv und übersichtlich.
5. Die Grafik ist gut strukturiert und leicht zu verstehen, ohne verwirrende Elemente.

Test 2:

Frage 1: Was fällt dir auf den ersten Blick auf?

1. Die Farbunterschiede stechen sofort ins Auge, besonders bei den Teams mit vielen Gelben Karten wie z.B. Leeds United.
2. Es ist auffällig, dass die Teams mit den meisten Punkten im oberen Bereich der Grafik weniger Gelbe Karten und Fouls haben.

3. Es gibt eine interessante Streuung der Teams, aber die Farbübergänge zwischen den Gelben Karten sind schwer zu unterscheiden.
4. Die Verteilung der Fouls ist sehr sichtbar, vor allem, dass die Top-Teams weniger Fouls haben.
5. Auf den ersten Blick sieht die Grafik etwas chaotisch aus, da die Farbskala für Gelbe Karten und die Verteilung der Punkte nicht sofort zusammenpassen.

Frage 2: Wie würdest du das Hauptthema oder die Hauptaussage dieser Visualisierung beschreiben?

1. Sie zeigt, wie Fouls und Gelbe Karten mit der Punktzahl eines Teams in der Premier League zusammenhängen.
2. Diese Grafik veranschaulicht den Zusammenhang zwischen Fouls, Gelben Karten und den Liga-Punkten der Teams.
3. Es wird versucht zu zeigen, dass Teams mit weniger Fouls und Gelben Karten erfolgreicher sind, aber es ist nicht sofort klar ersichtlich.
4. Es wird der Zusammenhang zwischen der Anzahl an Fouls, Gelben Karten und den erreichten Punkten in der Liga dargestellt.
5. Die Grafik will den Einfluss von Fouls und Gelben Karten auf die Punkte der Teams zeigen, wirkt aber etwas überladen.

Frage 3: Wie leicht fällt es dir, die Informationen aus dieser Visualisierung zu vergleichen und zu interpretieren? Gibt es Bereiche, die verwirrend oder unklar sind?

1. Es ist gut verständlich, aber die Farbunterschiede sind schwer zu erkennen, wenn Teams nah beieinander liegen.
2. Die Visualisierung ist insgesamt einfach zu verstehen, aber die Farbübergänge könnten klarer sein.
3. Einige Bereiche, besonders wo die Punkte dicht beieinander sind, sind verwirrend. Die Farbskala könnte besser sein.
4. Es ist gut zu vergleichen, obwohl die Position mancher Teams etwas nah ist und die Farbtöne sich zu wenig unterscheiden.
5. Ich finde die Grafik eher verwirrend, da die Farben und Positionen der Teams nicht immer ein klares Muster ergeben. Es könnte etwas übersichtlicher sein.

Test 3:

Frage 1: Was fällt dir auf den ersten Blick auf?

1. Die Symbole für Heim- und Auswärtsspiele stechen sofort ins Auge, zusammen mit den vielen Farbmarkierungen für die Ergebnisse.
2. Die Schwankungen im Ballbesitz sind deutlich, vor allem die grossen Unterschiede zwischen den Spielen.
3. Die durchgehende Linie verbindet alle Punkte, aber die Vielzahl an Informationen auf einmal ist etwas überwältigend.
4. Mir fällt auf, dass die Heimspiele oft einen höheren Ballbesitz haben, dargestellt durch die Dreiecke.

5. Auf den ersten Blick sind die Grössen der Punkte nicht so auffällig – die Tackling Erfolgsrate geht fast unter im Vergleich zu den anderen Variablen.

Frage 2: Wie würdest du das Hauptthema oder die Hauptaussage dieser Visualisierung beschreiben?

1. Sie zeigt, wie der Ballbesitz von Arsenal im Verlauf der Saison variiert, getrennt nach Heim- und Auswärtsspielen.
2. Es geht darum, den Ballbesitz von Arsenal über die Zeit darzustellen, wobei die Spiele nach Ergebnis unterschieden werden.
3. Die Grafik stellt den Ballbesitz dar, verknüpft mit den Ergebnissen und ob es ein Heim- oder Auswärtsspiel war.
4. Der Fokus liegt auf dem Zusammenhang zwischen Ballbesitz und den Spielergebnissen, besonders im Hinblick auf Heim- und Auswärtsspiele.
5. Zusätzlich zum Ballbesitz wird auch die Tackling Erfolgsrate integriert, aber sie bleibt eher im Hintergrund.

Frage 3: Wie leicht fällt es dir, die Informationen aus dieser Visualisierung zu vergleichen und zu interpretieren? Gibt es Bereiche, die verwirrend oder unklar sind?

1. Die Visualisierung ist insgesamt klar, aber mit vielen Dimensionen auf einmal vielleicht etwas zu komplex.
2. Die Unterscheidung der Spieltypen und Ergebnisse funktioniert gut, aber ich musste die Legende erst genauer anschauen.
3. Die vielen Farben und Formen sind etwas verwirrend, vor allem, wenn man versucht, alles gleichzeitig zu verstehen.
4. Die Darstellung ist übersichtlich, aber die Informationen könnten aufgeteilt werden, um die Lesbarkeit zu erhöhen.
5. Die Tackling Erfolgsrate ist gut integriert, aber die Punktgrössen sind nicht sofort ersichtlich, sie fallen im Vergleich zur Farb- und Formkodierung kaum auf.

Anhang

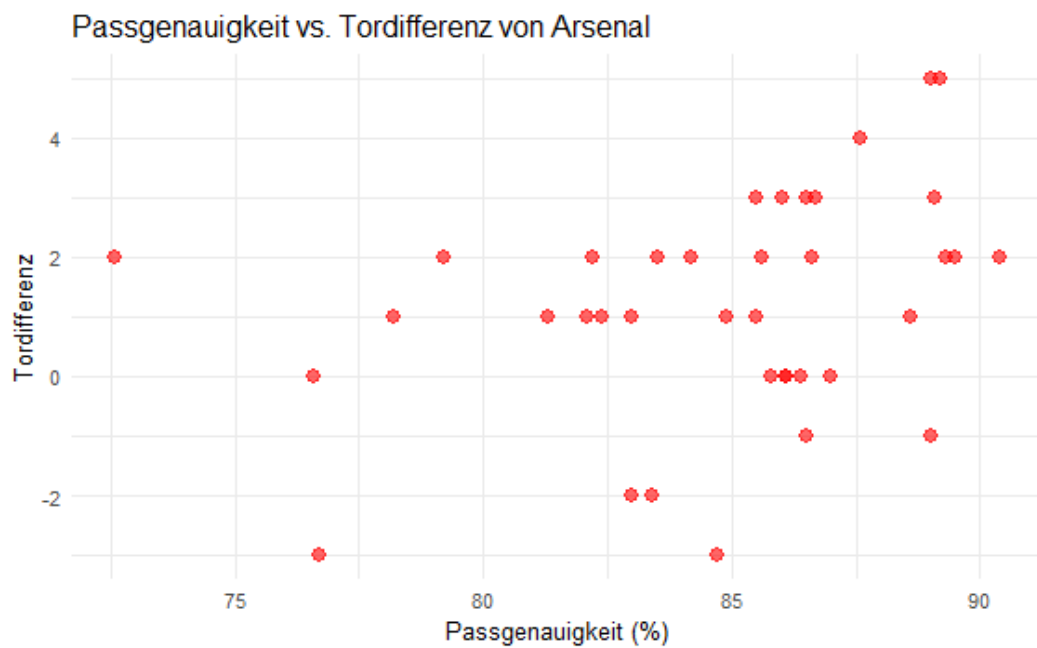


Abbildung 3: Passgenauigkeit vs. Tordifferenz von Arsenal

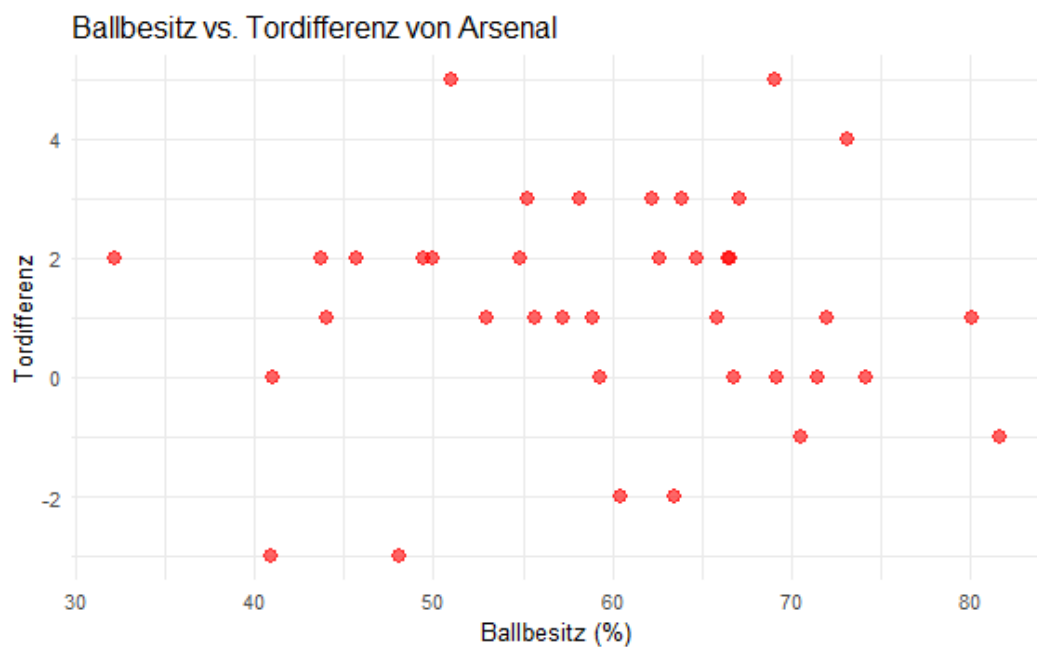


Abbildung 4: Ballbesitz vs. Tordifferenz von Arsenal

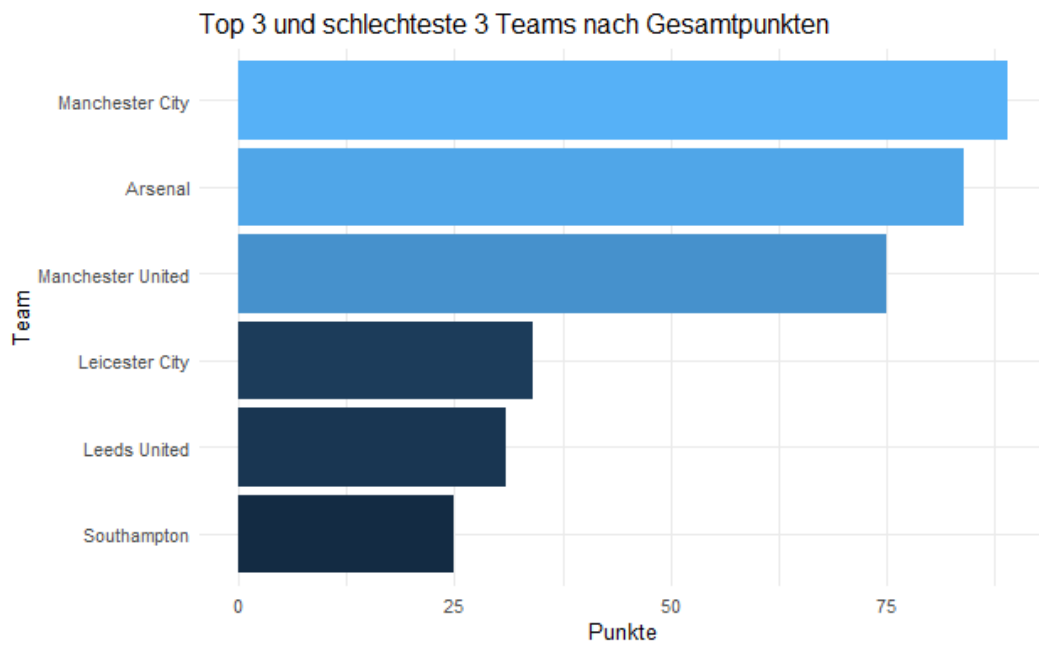


Abbildung 7: Top 3 und schlechteste 3 Teams nach Gesamtpunkten

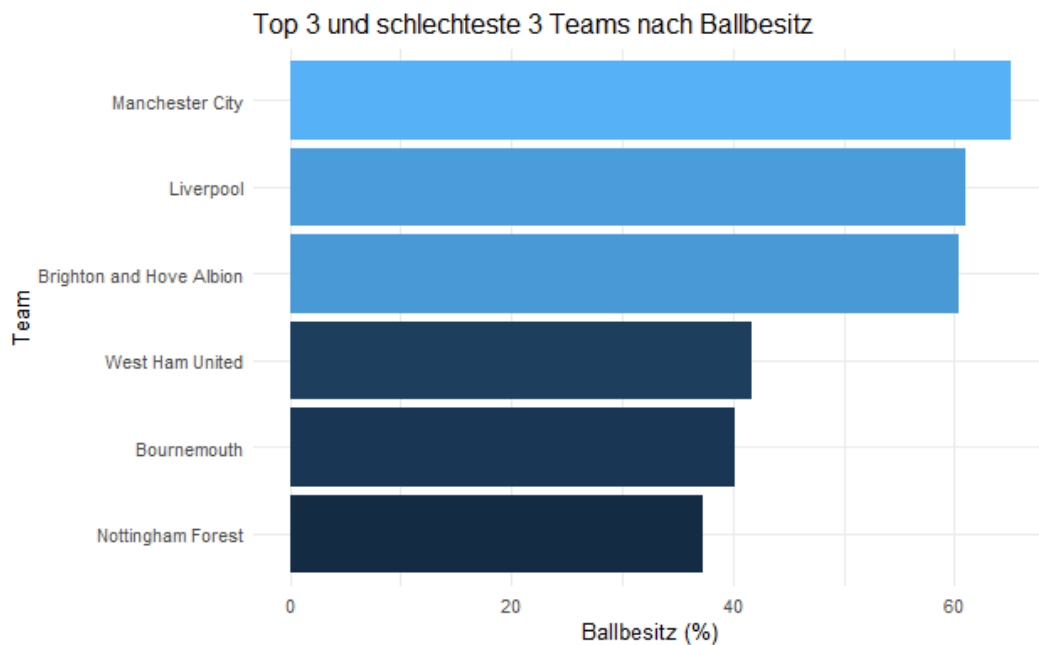


Abbildung 8: Top 3 und schlechteste 3 Teams nach Ballbesitz

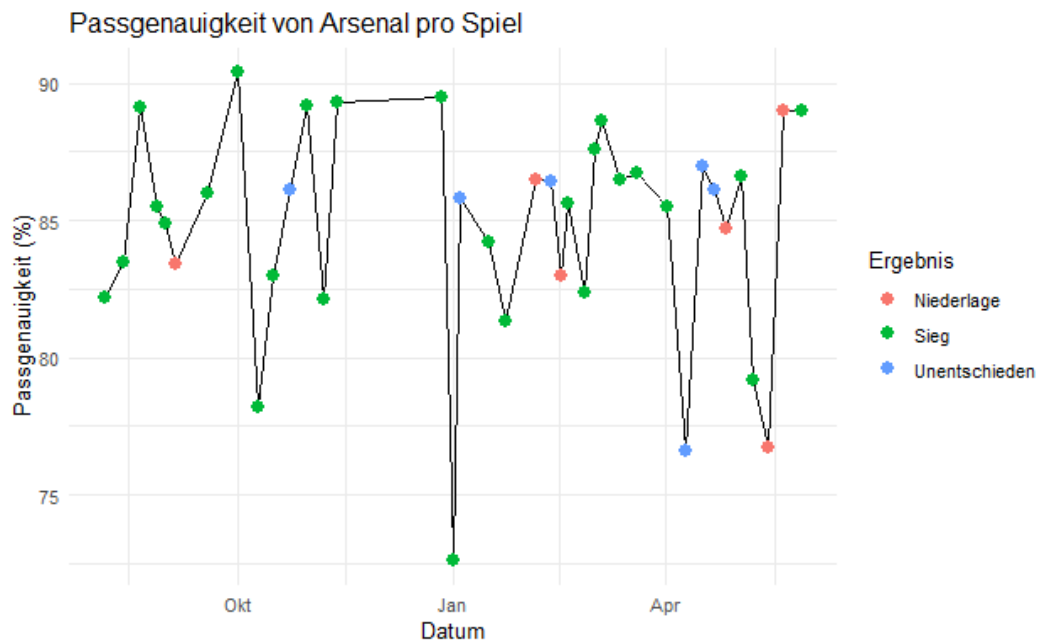


Abbildung 17: Passgenauigkeit von Arsenal im Verlauf der Saison

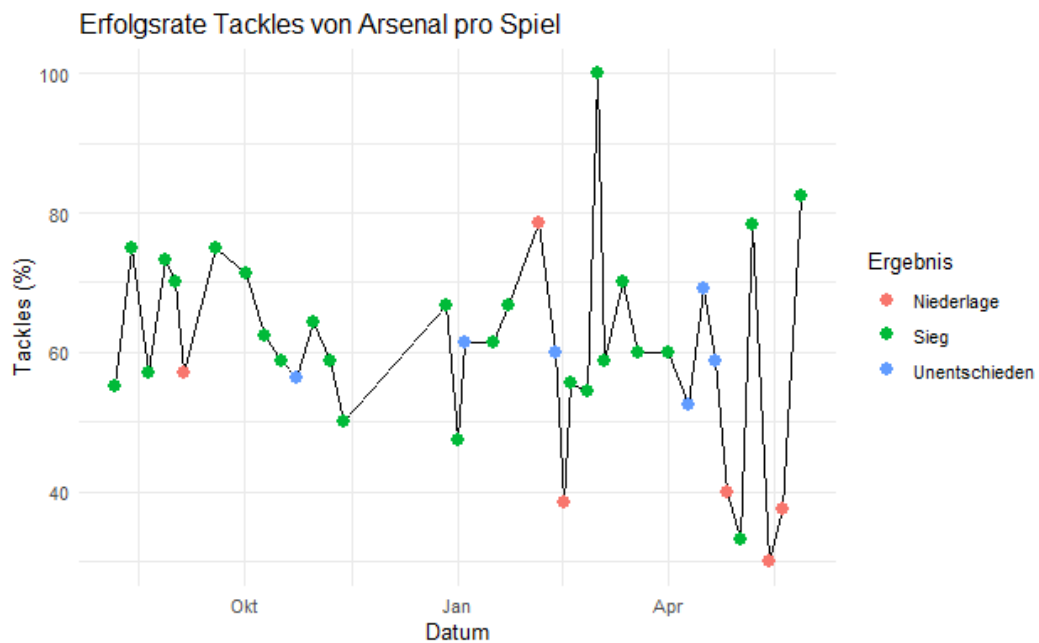


Abbildung 10: Erfolgsrate Tacklings von Arsenal im Verlauf der Saison

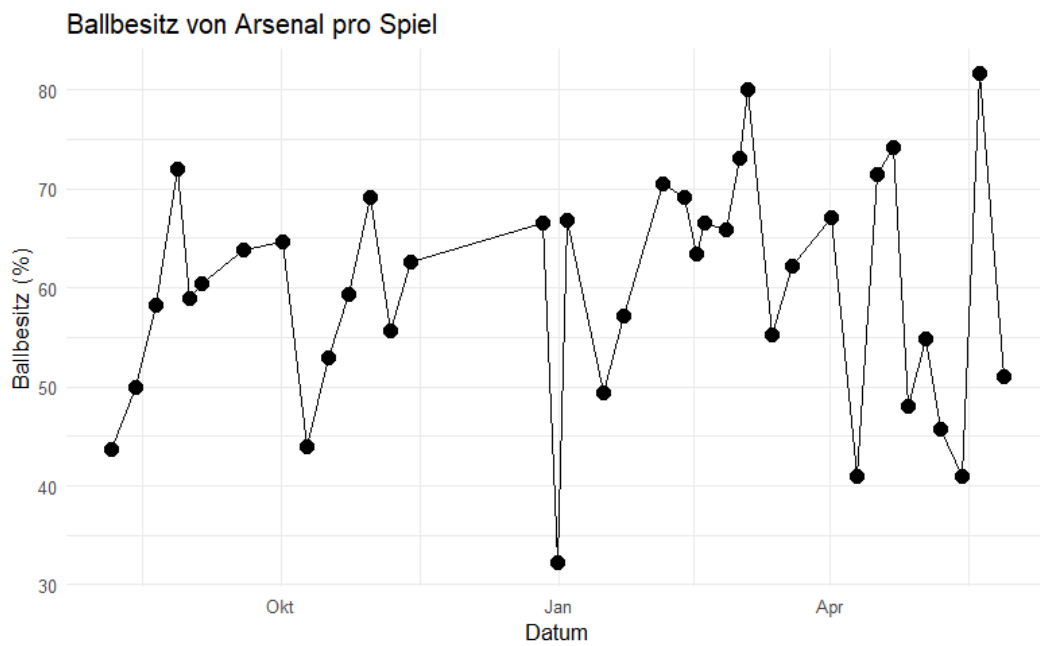


Abbildung 18: Ballbesitz von Arsenal im Verlauf der Saison (2-dimensional)

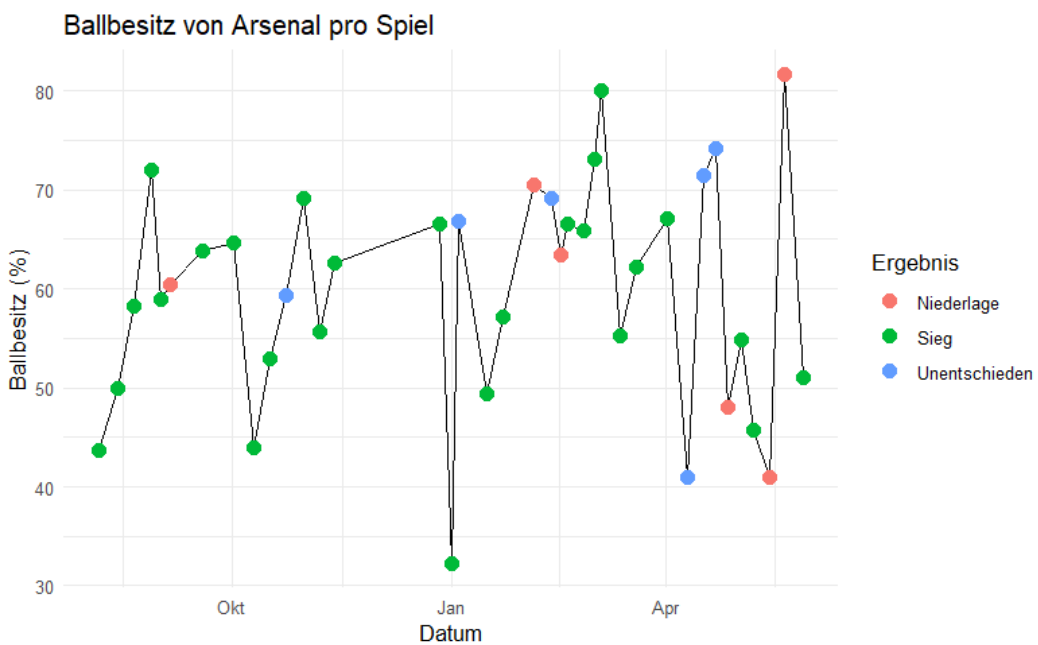


Abbildung 20: Ballbesitz von Arsenal im Verlauf der Saison (3-dimensional)

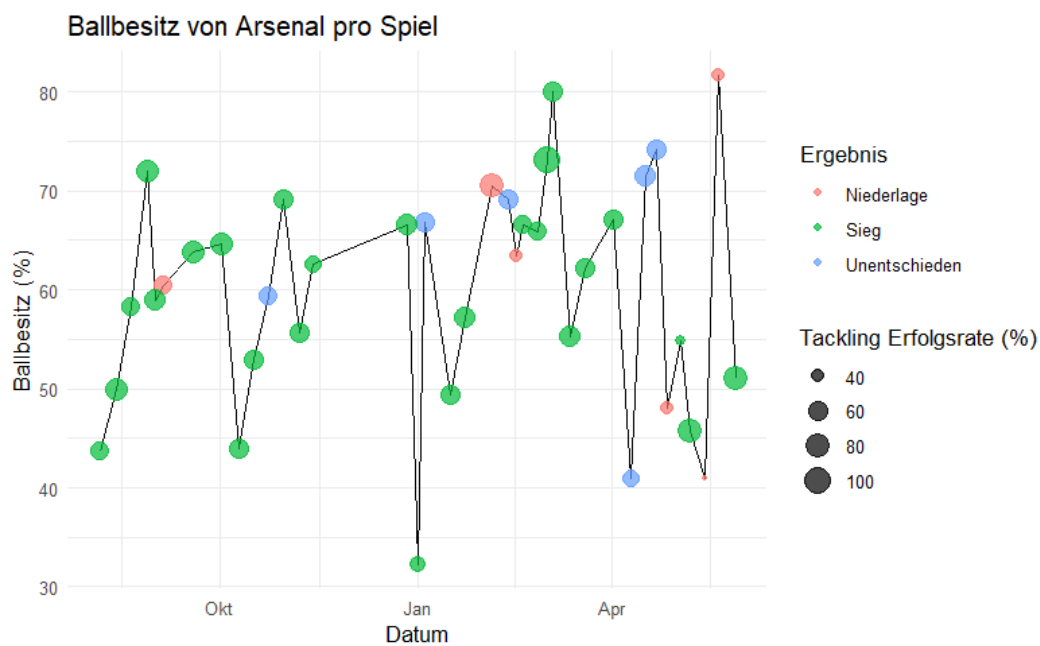


Abbildung 23: Ballbesitz von Arsenal im Verlauf der Saison (4-dimensional)

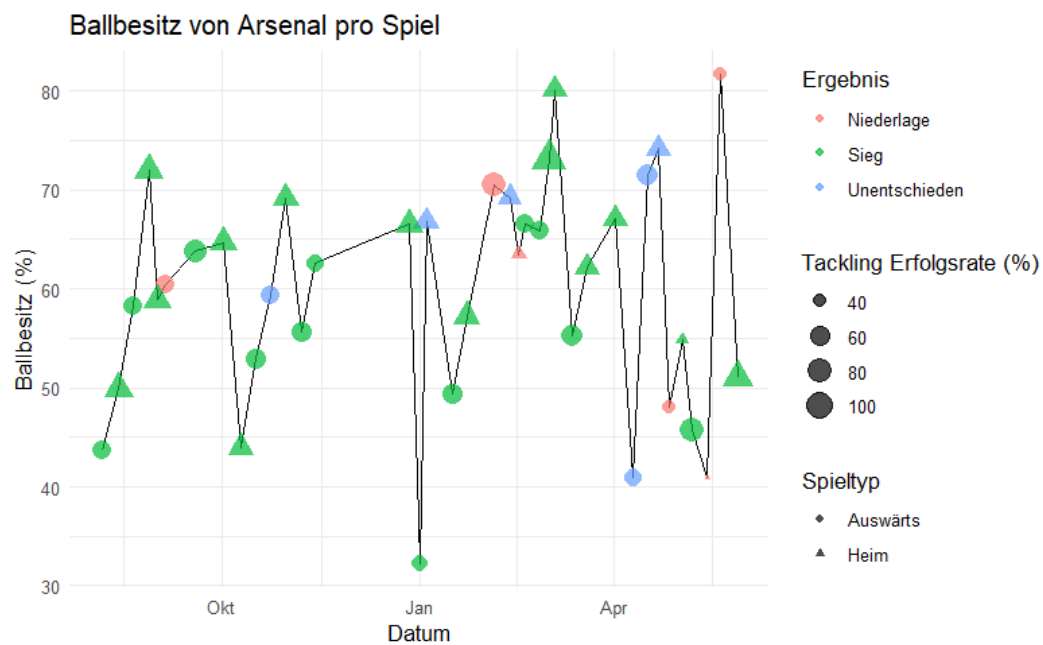


Abbildung 21: Ballbesitz von Arsenal im Verlauf der Saison (5-dimensional)

Heimteam	Durchschnitt Fouls	Durchschnitt Gelbe Karten	Punkte
Manchester City	9.13	1.16	89
Arsenal	9.82	1.37	84
Manchester United	11.16	2.05	75
Newcastle United	10.68	1.63	71
Liverpool	10.66	1.50	67
Brighton and Hove Albion	11.24	1.55	62
Aston Villa	10.97	2.11	61
Tottenham Hotspur	11.16	1.97	60
Brentford	9.32	1.45	59
Fulham	10.82	2.11	52
Crystal Palace	11.74	2.08	45
Chelsea	10.45	2.03	44
Wolverhampton Wanderers	12.18	2.21	41
West Ham United	9.53	1.16	40
Bournemouth	10.29	1.82	39
Nottingham Forest	11.71	2.21	38
Everton	10.42	2.08	36
Leicester City	10.79	1.55	34
Leeds United	12.29	2.21	31
Southampton	11.32	1.92	25

Abbildung 15: Durchschnittliche Fouls, durchschnittliche gelbe Karten und Punkte pro Team tabellarisch dargestellt

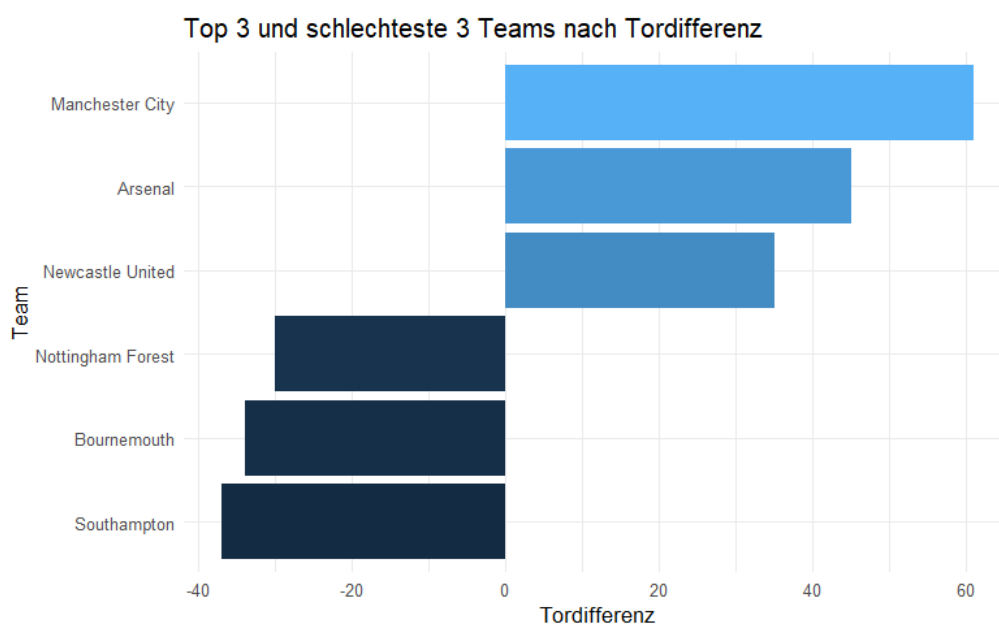


Abbildung 25: Top 3 und schlechteste 3 Teams nach Tordifferenz

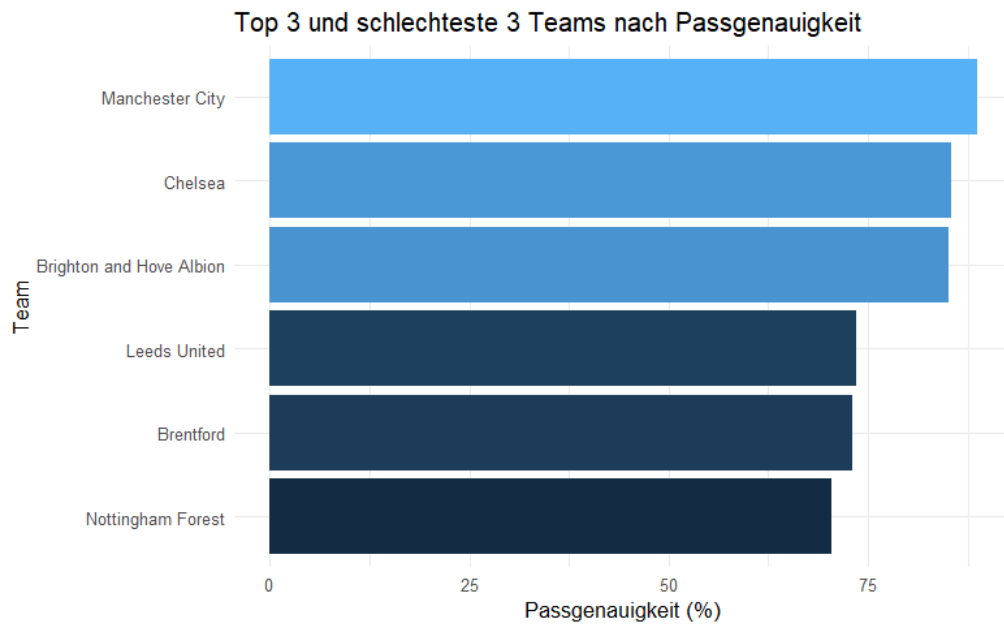


Abbildung 17: Top 3 und schlechteste 3 Teams nach Passgenauigkeit

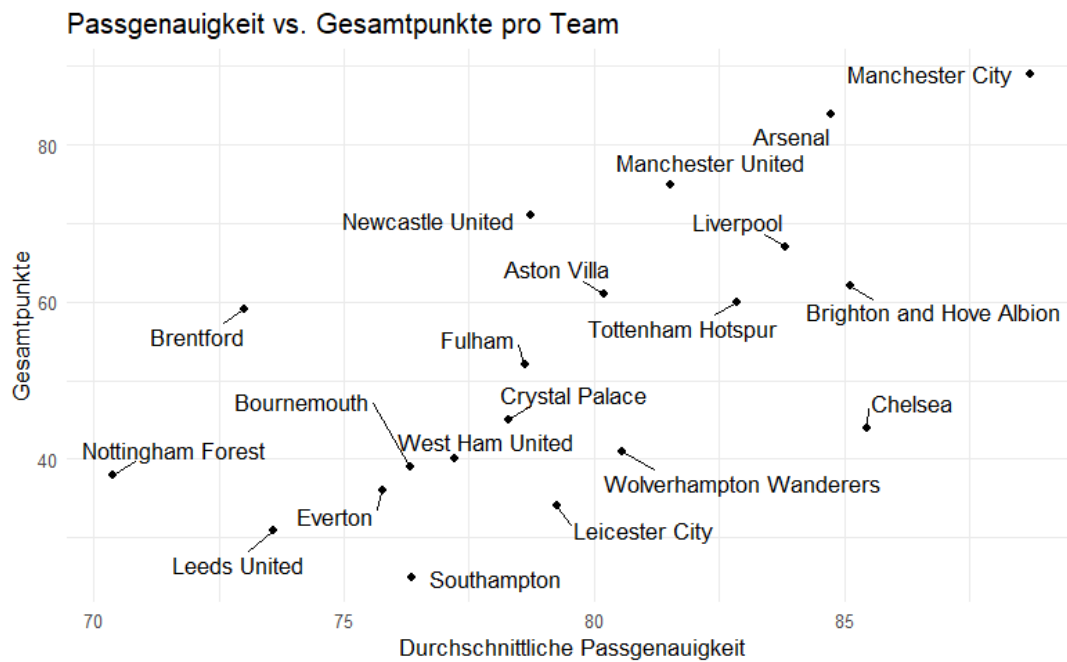


Abbildung 26: Passgenauigkeit vs. Gesamtpunkte pro Team