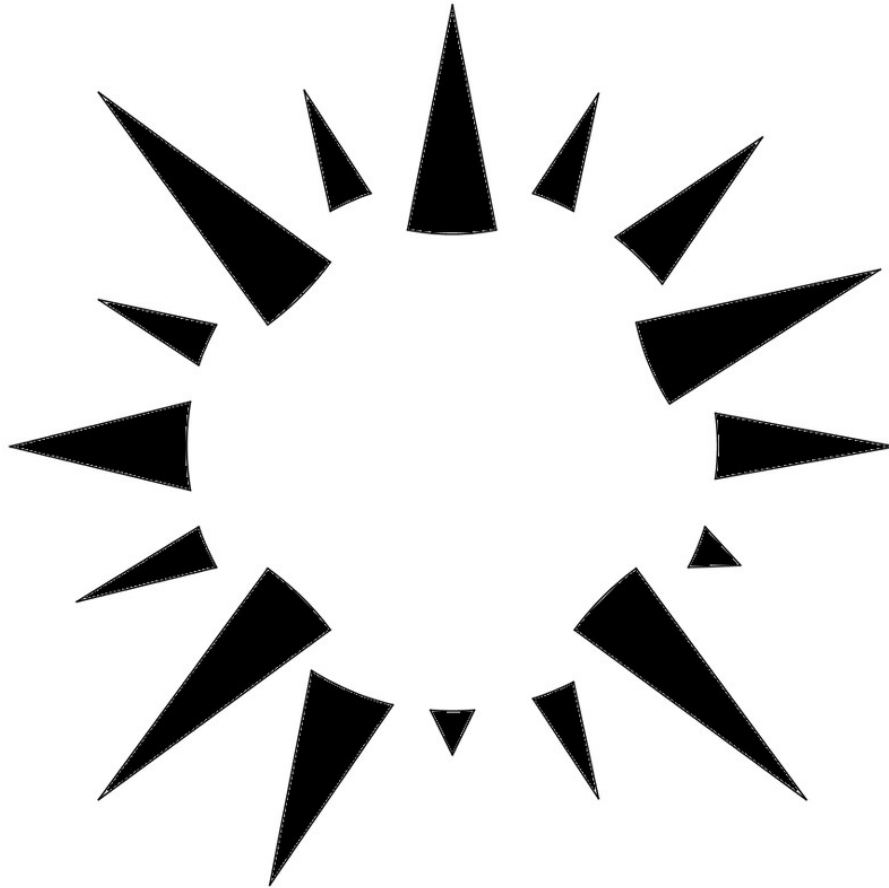


Grundkompetenz Datenvisualisierung



(„Scheinkonturen“, 2022)

Student

Si Ben Tran

Expertin

Prof. Dr. Arzu Çöltekin

GitHub

https://github.com/7ben18/grundlagen_datenvisualisierung

Fachhochschule Nordwestschweiz, Hochschule für Technik

Windisch, 24. Mai 2022

Inhaltsverzeichnis

1	LO1: Visualization basics, chart types	3
1.1	Streudiagramme	3
1.2	Blasendiagramm	3
1.3	Liniendiagramm	3
1.4	Histogramme	4
2	LO2: Visual Perception	5
2.1	Anpassung der Grösse	5
2.2	Anpassung der Orientierung	5
2.3	Anpassung der Formen	6
2.4	Anpassung der Transparenz	6
2.5	Anpassung der Farben	6
3	LO3: Design Principles vs. Data	7
3.1	Datenformat	7
3.2	Gestaltungsprinzipien von Skalierungen	8
3.3	Gestaltungsprinzipien von Legenden	8
3.4	Gestaltungsprinzipien von Tabellen	8
4	LO4: Grammar of Graphics Tools	9
4.1	Daten	9
4.2	Ästhetik	9
4.3	Größe	9
4.4	Geometrische Objekte	9
4.5	Statistik	10
4.6	Facetten	10
4.7	Koordinaten System	10
5	LO5: Evaluation	11
5.1	Usability-Test	11
5.2	Durchführung der Umfrage	12
5.3	Auswertung der Umfrage	12
6	Anhang Abbildungen	13
7	Quellenverzeichnis	27
8	Abbildungsverzeichnis	29
9	Ehelichkeitserklärung	30

1 LO1: Visualization basics, chart types

Visualisierungen können sehr hilfreich sein, um Informationen und Daten zu erkennen und zu verstehen. In der Data Science Welt ist Datenvisualisierung ein essenziell wichtiges Verfahren, um grosse Mengen an Informationen zu analysieren (*Handbuch zur Datenvisualisierung*, 2022). In diesem Kapitel geht es darum, die grundlegenden Visualisierungsarten zu zeigen und was bei einer Visualisierung zu beachten ist. Die angesprochene Zielgruppe sind Personen, die sich für die verschiedenen Grundlagenvisualisierungsarten bei Gapminder Daten interessieren. Dabei werden wir den Datensatz gapminder von der R Bibliothek gapminder nutzen. Gapminder ist ein Datensatz der uns für jedes Land die Informationen über Lebenserwartung der Menschen, die Gesamtpopulation und Bruttoinlandsprodukt, unterteilt in Länder und Kontinente jeweils in einem Zeitintervall in 5 Jahren von 1952 bis 2007 beinhaltet. (Bryan & MacDonald, 2015)

1.1 Streudiagramme

Ein Streudiagramm, auch Punktwolke genannt, ist eine graphische Darstellung von Wertepaaren. Die Wertepaare werden in einem kartesischen Diagramm entsprechend dem x und y-Wert als Punkt aufgetragen (Broese et al., 2020) Abbildung 2. Ein Streudiagramm kommt an seine Grenze, sobald zu viele Observationen vorhanden sind und mehrere Punkte sich überlappen. Man kann die Sättigung der Punkte minimieren, die Punkte verkleinern oder in unterschiedlichen Kategorien aufteilen oder ein 2-d Histogramm (Heatmap) erstellen um eine bessere Übersicht zu erhalten (Yi, 2021). Empfehlenswert ist zudem das Hinzufügen von Farben, aufgrund einer kategorischen Variabel, um mehr Informationen zu übermitteln Abbildung 3. Bei Abbildung 2 bis 4, ist noch zu beachten, dass die y-Achse bei der Skalierung 30 Jahren startet.

1.2 Blasendiagramm

Ein Blasendiagramm ist ein Streudiagramm, bei der eine weitere Dimension hinzugefügt, die durch die Grösse der Blasen dargestellt wird. Somit ist die Blase abhängig von einer Quantitativen variabel (*Streu- und Blasendiagramme - Hilfe für Data Studio*, o. J.). Beim Gapminder Datensatz, wird oft für die Lebenserwartung und BIP ein Blasendiagramm erstellt. Dabei sind die Achsenbeschriftungen sowie Titel und Legenden essenziell wichtig, um das komplette Wissen aus der Visualisierung herauszuschöpfen zu können Abbildung 4.

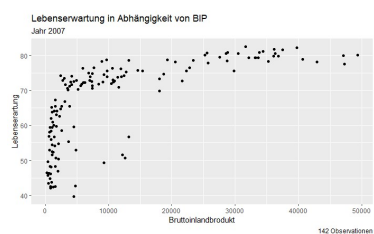


Abbildung 2 Streudiagramm

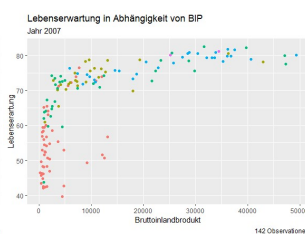


Abbildung 3 Streudiagramm mit Kategorische variabel

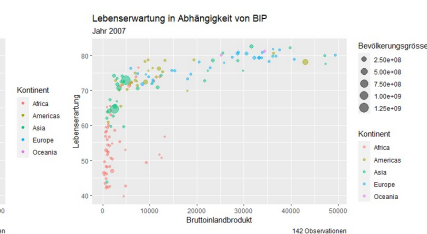


Abbildung 4 Blasendiagramm

1.3 Liniendiagramm

Verbindet man in einem Streudiagramm die Punkte miteinander, erhält man ein Liniendiagramm, auch Kurvendiagramm genannt (*Kurven- und Punktdiagramme einfach erklärt mit unserer Hilfe*, o. J.) Abbildung 5. Ein Liniendiagramm wird oft verwendet, um über einen bestimmten Zeitintervall eine Entwicklung, Veränderung und Trends zu visualisieren oder festzustellen (*Wie findet man das richtige Diagramm für seine Daten?*, o. J.). In Abbildung 5 wurden die Daten nach der durchschnittlichen Lebenserwartung in Jahren und Jahr aggregiert. Wobei bei der Aggregation eine Datenreduktion stattfindet, welches bei einem Streu- oder Blasendiagramm nicht der Fall war und somit zusätzliche Informationen verloren gehen (Jugel et al., 2014, S. 800). Um nun diese Datenreduktion zu minimieren, kann man zusätzlich eine weitere kategorische Variabel aufnehmen und diese dann visualisieren Abbildung 6.

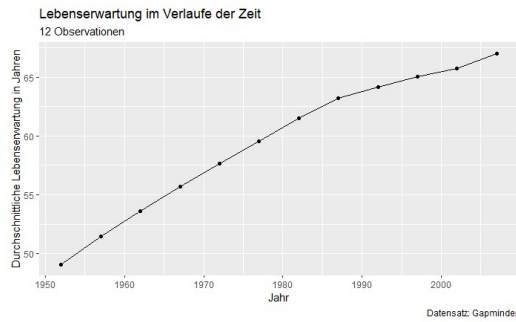


Abbildung 5 Liniendiagramm

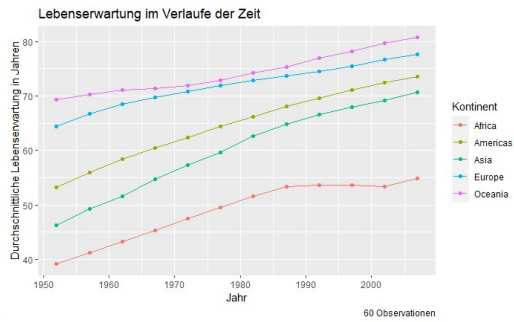


Abbildung 6 Liniendiagramm mit Kategorische Variabel

1.4 Histogramme

Histogramm ist eine Visualisierung, welches die Merkmalsausprägungen der Daten versucht darzustellen. Oft wird ein Histogramm erstellt, um eine Verteilung der quantitativen Daten besser zu verstehen. Bei einem Histogramm wird die Anzahl vorkommenden Observationen gezählt und mittels Bildung von „Bins“, auch bekannt als Anzahl Klassenbreite, diskretisiert (Perruchoud, 2022, S. 24, 28). Somit ist die Visualisierung eines Histogramms von der Anzahl der Klassenbreite abhängig. Bei einer kleinen Anzahl der Klassenbreite werden die Merkmale der Verteilung nicht angezeigt und gehen somit verloren Abbildung 7. Auf der anderen Seite wird bei einer grossen Anzahl Klassenbreite das Histogramm unübersichtlich und die Haupttrends der Daten können verloren gehen Abbildung 8. Bei einem Histogramm muss man verschiedene Anzahl Klassenbreite überprüfen, um eine optimale Visualisierung zu finden, welches die Daten widerspiegelt (Wilke, 2019, S. 59–61) Abbildung 9.

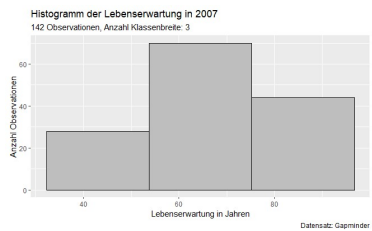


Abbildung 7 Anzahl bins = 3

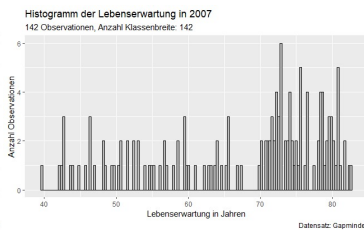


Abbildung 8 Anzahl bins = 142

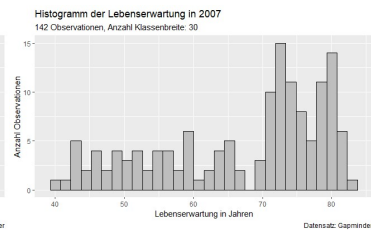


Abbildung 9 Anzahl bins = 30

Ein Histogramm kann in qualitativen Daten unterteilt werden. Dabei zu beachten ist, dass alle kategorischen Daten unterschiedliche Farbe haben und voneinander klar trennbar sind (Wilke, 2019, S. 65–66). In Abbildung 10 sind die Histogramme aufeinander gestapelt. Das Interpretieren solcher Histogramme ist sehr schwierig. Einfacher sieht es bei der Abbildung 11 aus. In dieser Visualisierung ist das Interpretieren einfacher, weil die Ausprägung entsprechend kategorischen Daten und Überlappung visualisiert sind. Bei der Abbildung 12 ist die Überlappung nicht vorhanden, jedoch die Ausprägungen der einzelnen kategorischen Werte.

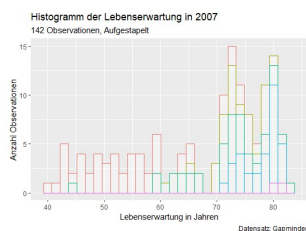


Abbildung 10 Aufstaplung

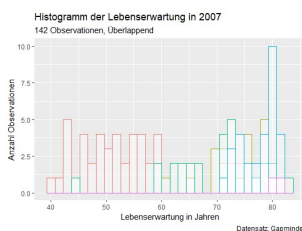


Abbildung 11 Überlappung

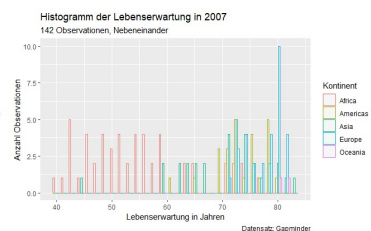


Abbildung 12 Nebeneinander

2 LO2: Visual Perception

Die Wahrnehmung ist subjektiv abhängig, jeder Mensch nimmt seine Umwelt und Visualisierungen anders wahr. Faktoren, wie das Geschlecht, Alter, Bildungsniveau und vieles mehr beeinflussen die Wahrnehmung. Neben der subjektiven Wahrnehmung ist der Mensch fähig, Visualisierungen aus einer objektiven Betrachtungsweise zu beurteilen (Clemens, 2016). In diesem Kapitel werden wir anhand der Abbildung 13 herausfinden, welche unterschiedliche Variablen einer Visualisierung unterschiedliche Wahrnehmungen auf den Menschen auslösen. Dabei gehen wir immer von der ursprünglichen Abbildung 13 aus und ändern dort eine Variabel. Konkret beziehen wir uns auf einen Pokémon Datensatz. Die Daten beinhalten verschiedene kategorische und quantitative Daten. Um eine Überladung der Streudiagramme zu verhindern, wurde der Datensatz nach dem Typ 1, Feuer gefiltert. Auf der x-Achse befinden sich die totalen Statik und auf der y-Achse wurde das Verhältnis zwischen Angriffswerten und Defensivwerten der Pokémon berechnet (Rounak, 2017).

2.1 Anpassung der Grösse

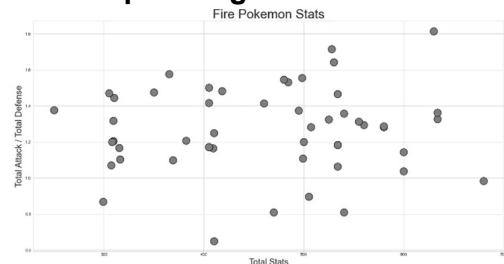


Abbildung 13 Grafik vor Anpassung

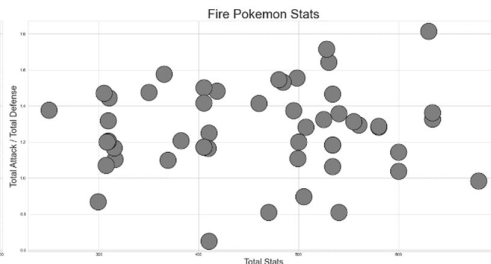


Abbildung 14 Grösse

Die Abbildung 13 ist ein Streudiagramm, bei welchem die Wertepaare entsprechend dem x- und y-Wert auf ein kartesisches Koordinatensystem geplottet wurde. Diese Visualisierung wurde bearbeitet, nämlich wurden die Grösse der Streupunkte geändert Abbildung 14. In dieser Visualisierung erkennt man einen deutlichen Unterschied zur Vorherigen. Die Visualisierung der Punkte wird stärker wahrgenommen und das Interpretieren der Wertepaare fällt einem schwerer, da sich die Punkte überlappen, was in der vorherigen Visualisierung nicht der Fall war. (Ware, 2013, S. 205–206).

2.2 Anpassung der Orientierung

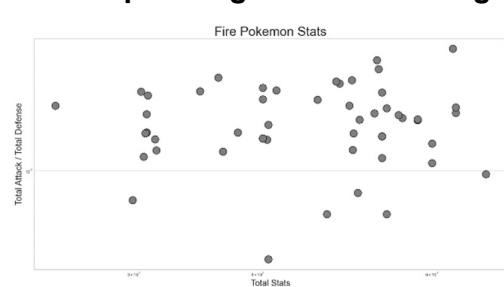


Abbildung 15 Skalen Logarithmieren

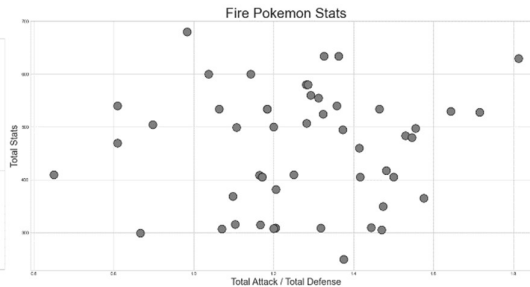


Abbildung 16 x- und y-Achse vertauschen

In der Abbildung 15 wurde die Skalierung logarithmisch angepasst, dies führt zu einer «Verschiebung» der ursprünglichen Visualisierung. Die logarithmische Skalierung kann von Nutzen sein, wenn viele Punkte sich in unmittelbarer Nähe zueinander befinden (Wilke, 2019, S. 17–18). Neben der logarithmischen Skalierung wurden in Abbildung 16 die Achsen vertauscht. Da es sich um ein Streudiagramm handelt, ist der visuelle Effekt nicht stark vorhanden, verglichen bei einem Liniendiagramm oder Balkendiagramm deutlicher.

2.3 Anpassung der Formen

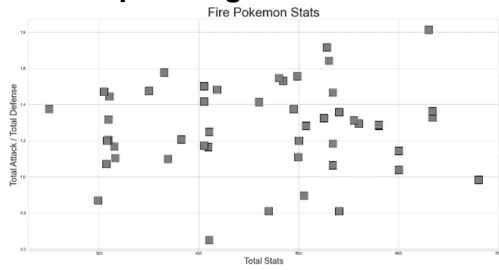


Abbildung 17 Quadratische Formen

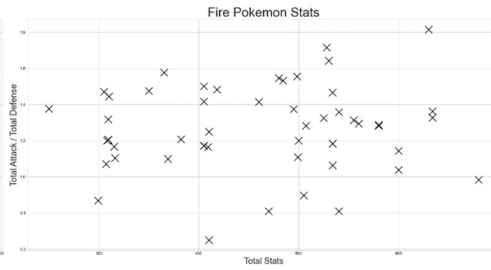


Abbildung 18 Kreuz Formen

In der Abbildung 17, 18 wurden die Formen der Streupunkte verändert. Bei der Abbildung 16, befinden sich nun Kreuze, mit diesen Kreuzen lassen sich nun die Wertepaare präziser zuordnen und man kann diese genauer ablesen. Durch die Schnittpunkte der beiden diagonalen Linien der Kreuze, ist der Interpretationsspielraum nicht mehr gegeben, ob nun der tatsächliche Wert sich bei der Abbildung 17 nun am linken oder rechten Rand befindet.

2.4 Anpassung der Transparenz

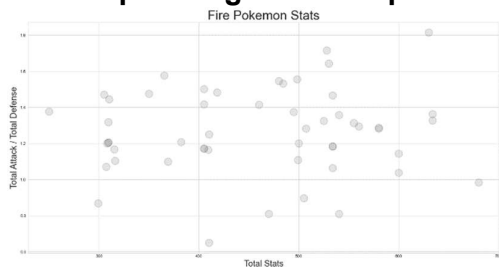


Abbildung 19 Transparenz

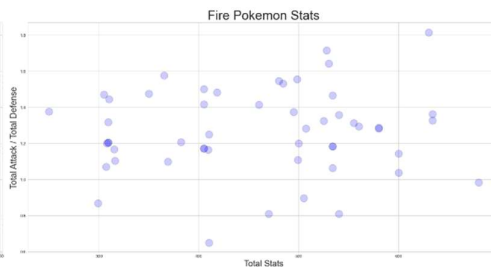


Abbildung 20 Transparenz mit Blau

In der Abbildung 19, 20 wurde die Transparenz der Streupunkte herabgesetzt. Dadurch erkennen wir, die überlappenden Punkte deutlich besser. Diese erscheinen im Streudiagramm intensiver (Ware, 2013, S. 205–206). Dadurch, dass Grau einen ähnlichen Kontrast zum weissen Hintergrund hat, wurden die Punkte zusätzlich blau eingefärbt, um diese vom Hintergrund hervorheben zu lassen.

2.5 Anpassung der Farben

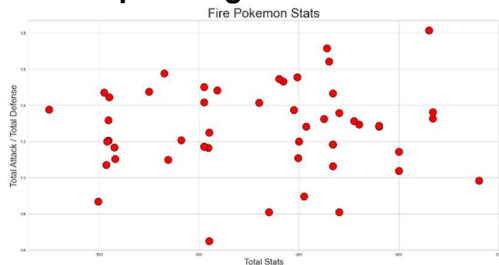


Abbildung 21 eingefärbt mit Rot

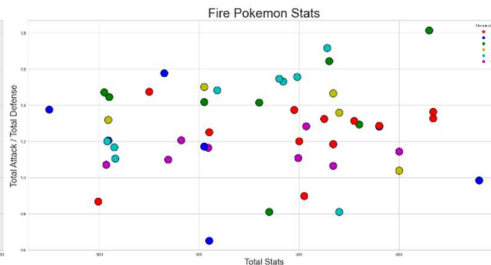


Abbildung 22 eingefärbt aufgrund kategorische Variabel

Neben der Grösse, Formen und Orientierung wurde die Wahrnehmung der Farben nur kurz in der Abbildung 20 erwähnt. In Abbildung 21 nehmen wir die roten Streupunkte sehr stark wahr verglichen zu den ursprünglichen grauen Punkten (Ware, 2013, S. 124). Dies hat damit zu tun, dass die Farbe vom weissen Hintergrund hervorgehoben wird. Auch ist die Farbe Rot aufmerksamkeitsstark, plakativ, stimulierend, wärmend sowie anregend. Jedoch kann je nach Wahrnehmung die Farbe Rot als Gefahr, Aggressivität, Zerstörung und Wut wahrgenommen werden (*Bedeutung der Farben* ["Farbsymbolik und Farben in der Psychologie"], 2020). Neben das Hervorheben der Punkte, sind die Farben auch geeignet, um weitere Informationen, wie der einer kategorischen Variabel aufzunehmen und durch geeignete Visualisierung zu vermitteln (Anton, 2019, Kapitel. 9.2.3). Diese wird in der Abbildung 22, nach Pokémon Generation als Beispiel umgesetzt.

3 LO3: Design Principles vs. Data

In diesem Kapitel behandeln wir, wie die Daten verarbeitet und strukturiert werden müssen, um eine geeignete Visualisierung zu erhalten. Zusätzlich befassen wir uns mit den verschiedenen Gestaltungsprinzipien von Grafiken und Tabellen, dabei nutzen wir den Datensatz Gapminder von LE1.

3.1 Datenformat

Der Datensatz wird in R Studio eingelesen und ausgegeben. Bei einem unstrukturierten Datensatz lohnt es sich gegeben falls, den Datensatz zu bereinigen, sprich die fehlende Werte zu imputieren oder nach einer zu sortieren. Um nun eine geeignete Visualisierung zu finden, müssen wir uns bewusst sein, mit welchen Datentypen wir arbeiten. Datentypen können in zwei Kategorien eingeteilt werden, nämlich in die qualitativen und quantitativen Daten. Diese wiederum können nochmals unterteilt werden. Zu den qualitativen Daten gehören die nominal und ordinalen Daten. Bei den quantitativen Daten sind es die diskreten und stetigen Daten (*Cramer und Kamps - 2020 - Grundlagen der Wahrscheinlichkeitsrechnung und Sta.pdf*, o. J., S. 6). Im vorhandenen Datensatz ist ein ordinaler qualitative (Jahr) sowie mehrere stetige quantitative (lifeExp, pop, gdpPercap) Daten vorhanden. Der Kontinent und das Land sind nominale Daten.

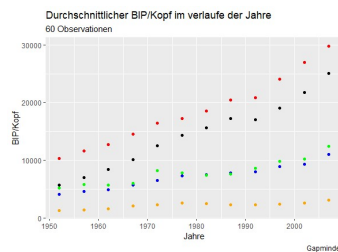


Abbildung 23 ø Punktdiagramm

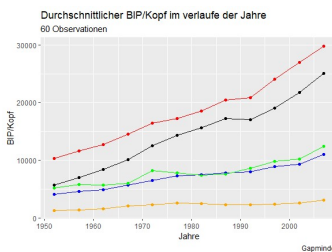


Abbildung 24 ø Liniendiagramm

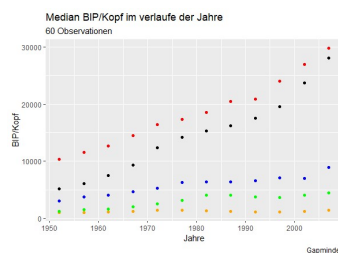


Abbildung 25 Median Punktdiagramm

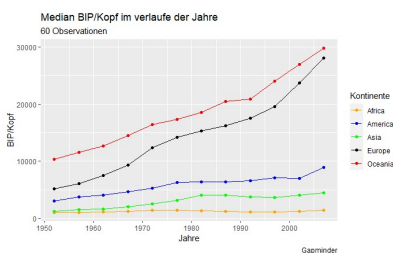


Abbildung 26 Median Liniendiagramm

Auch muss man sich im Vorhinein bewusst sein, welche Visualisierungsart und Aggregation man braucht. Vergleicht man Abbildung 23 & 24 untereinander erkennt man, dass bei Abbildung 24 Datenpunkte für bestimmte Jahre fehlen und durch Interpolation der Punkte eine Kontinuität suggeriert wird. Vergleicht man die Abbildungen 23, 24 mit 25, 26, erkennen wir einen leichten Unterschied, zwischen Median und Mittelwert. Ein weiteres Gestaltungsprinzip, welches man berücksichtigen sollte, sind Visualisierungen die Farbenblind freundlich sind (Ferreira, 2021).

Gestaltungsprinzipien der y-Achse

Neben der Achsenbeschriftung ist es auch wichtig zu erkennen, mit welchen x und y-Achsen gestartet wird. Bei den ersten Abbildungen von LE1, Abbildung 2 bis 4 erkennen wir an der y-Achse, dass der Startwert bei 40 Jahren liegt. Den meisten ist es klar, dass eine durchschnittliche Lebenserwartung von unter 40 sehr unwahrscheinlich ist, trotzdem ist der Interpretationsspielraum offen und man weiss nicht, ob nun nach den Werten unter 40 gefiltert wurde oder nicht. Entsprechend sind Achsenskalierungen zu beschriften, damit der Leser sich zu 100 % sicher ist, wie man die Visualisierung interpretieren soll (Wilke, 2019, S. 208–217).

3.2 Gestaltungsprinzipien von Skalierungen

In Abbildung 27 und 28 sehen wir den BIP/Kopf Verlauf vom Kontinent Afrika mit unterschiedlichen y-Achsen Skalierungen. In Abbildung 27, wird einem suggeriert, dass der BIP/Kopf sich in Afrika nicht bzw. nur wenig geändert hat. Jedoch, wenn man die Abbildung 28 betrachtet, erkennen wir deutlich, dass der BIP/Kopf deutlich fast um das Dreifache gestiegen ist. Die richtige Skalierung der y-Achse ist wichtig, damit die Visualisierung nicht die wichtigen Informationen verlieren (Wilke, 2019, S. 209–211).

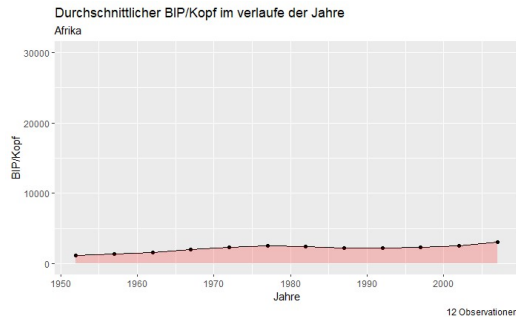


Abbildung 27 Skalierung der y-Achse (0 – 30000)

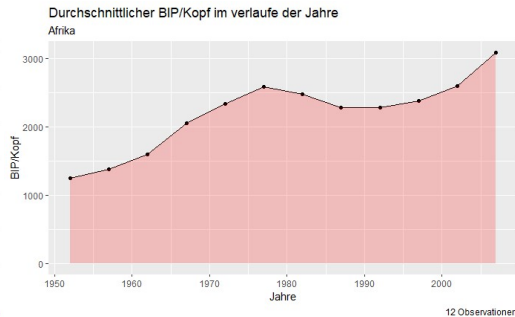


Abbildung 28 Skalierung der y-Achse (0 – 3000)

3.3 Gestaltungsprinzipien von Legenden

Einer der Gestaltungsprinzipien von Legenden, dass die Reihenfolge der Legenden den Daten entsprechend passen. Siehe die Abbildung 29 und Abbildung 30 als Vergleich. Auch ist es hilfreich für Personen mit Farbschwäche, komplementär Farben auszuwählen (Wilke, 2019, S. 248–250).

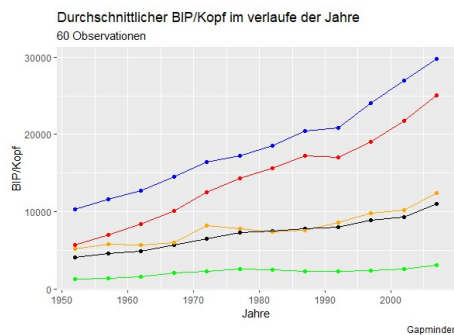


Abbildung 29 Legende ohne Reihenfolge

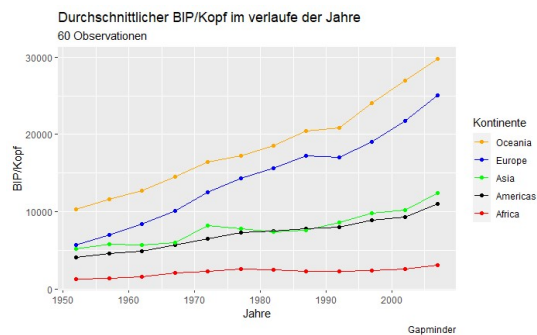


Abbildung 30 Legende mit Reihenfolge

3.4 Gestaltungsprinzipien von Tabellen

Neben den Visualisierungen sind Tabellen ebenfalls eine Art der Darstellung von Daten. Auch bei Tabellen gibt es einige Regeln, die man beachten sollte. In Abbildung 31 wird eine „hässliche“ Tabelle repräsentiert und in Tabelle 32 eine, die den Konventionen der Visualisierungen übereinstimmt. Tabellen benötigen keine Senkrechte Linien, Textwerte sollten linksbündig und Zahlen mit der gleichen Anzahl Kommastellen rechtsbündig sein. Zudem muss der Titel von den restlichen Daten visuell getrennt werden und man kann die Farben als effiziente Gruppierung der Daten innerhalb von Zeilen nutzen (Wilke, 2019, S. 273–275).

country	continent	year	lifeExp	pop	gdpPercap
Japan	Asia	2007	82.603	127467972	31656.0681
Hong Kong	Asia	2007	82.208	6980412	39724.9787
Iceland	Europe	2007	81.757	301931	36180.7892
Switzerland	Europe	2007	81.701	7554661	37506.4191
Australia	Oceania	2007	81.235	20434176	34435.3674

Abbildung 31 „Hässliche“ Tabelle

country	continent	year	lifeExp	pop	gdpPercap
Japan	Asia	2007	82.603	127467972	31656.0681
Hong Kong	Asia	2007	82.208	6980412	39724.9787
Iceland	Europe	2007	81.757	301931	36180.7892
Switzerland	Europe	2007	81.701	7554661	37506.4191
Australia	Oceania	2007	81.235	20434176	34435.3674

Abbildung 32 „Schöne“ Tabelle

4 LO4: Grammar of Graphics Tools

Wir gehen einen Schritt zurück, und befassen uns in diesem Abschnitt mit dem Thema «Grammar of Graphics Tools». Ein Konzept, das von Leland Wilkinson vorgeschlagen wird und alle wichtigen Aspekte einer effektiven Visualisierung im Detail abdeckt (Sarkar, 2018). Dieses Kapitel sollte eine Übersicht darstellen, wie man die Dimension einer Visualisierung Stück für Stück weiter erhöhen kann. Konkret werden wir uns wieder mit den Pokémon Daten widmen und uns in die Thematik Schritt für Schritt einarbeiten.

4.1 Daten

Zuerst starten wir mit den Daten, hier betrachten wir, mit welchen Daten wir arbeiten und was wir gerne visualisieren möchten. In der Abbildung 33 wird ein Streudiagramm gezeigt, welches von Feuer, Wasser und Gras Pokémon den Wertepaar-Angriff und Verteidigung mittels einem Punkteplot aufzeigt. Aufgrund der grossen Anzahl an Pokémon, die es gibt, wurden ca. 100 zufällig ausgewählt.

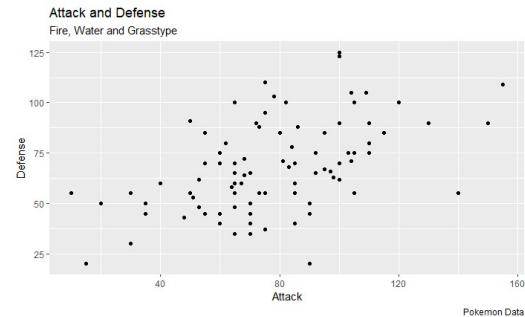


Abbildung 33 Zwei Dimensionen

4.2 Ästhetik

In der Ästhetik befassen wir uns, ob die x- und y-Achsen die richtigen Skalierungen aufweisen. Zudem überprüfen wir, ob wir durch weitere Variabel die Dimension der Visualisierung erhöhen können. In der Abbildung 34 erhöhen wir die Dimension, indem wir den Pokémon Typ visuell mit entsprechender Farbe einfärben.

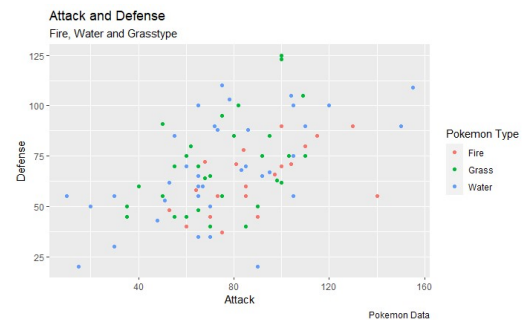


Abbildung 34 Drei Dimensionen

4.3 Größe

Die Grösse der Punkte erlaubt uns, eine weitere Dimension hinzuzufügen. Potenziell mehrere Werte zu skalieren oder einen Bereich dafür zu erstellen. In der Abbildung 35 wurden die totalen Status der Pokémon als Skalierungsgrösse angegeben und visuell durch die grösse der Punkte dargestellt.

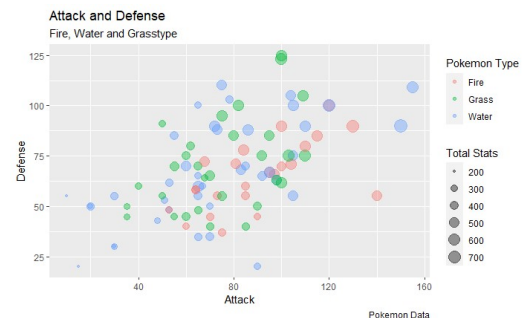


Abbildung 35 Vier Dimensionen

4.4 Geometrische Objekte

Mit den geometrischen Objekten können wir die Dimension weiter erhöhen, dabei können wir verschiedene Muster nutzen für verschiedene kategorischen Variabel. In der Abbildung 36 erkennt man aufgrund vom Muster, ob das Pokémon einen zweiten Typ neben Feuer, Wasser und Grass besitzt.

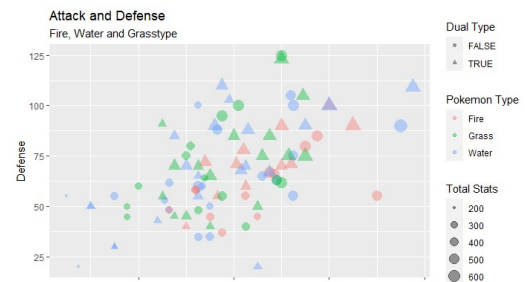


Abbildung 36 fünf Dimensionen

4.5 Statistik

Die Statistik erlaubt es uns, die Visualisierung auf eine weitere Dimension zu erhöhen. Dabei können Mittelwerte, Median oder Konfidenzintervalle mit visualisiert werden. Hier in der Abbildung 37 werden die Mittelwerte von der x- und y-Achsen eingezeichnet. Aufgrund der Einteilung in Quadranten können wir nun aussagen wie unterdurchschnittlich oder überdurchschnittlich treffen. Der obere rechte Quadrant teilt die Punkte in die überdurchschnittlich hohen Attack und Defense Werten ein, während das Gegenstück im unteren linken Quadranten sich befindet. Der Linke oberer Quadrant hat einen hohen durchschnittlichen Defense Wert jedoch tiefe Attack Werte. Das gegenteilige Quadrant befindet sich rechts unten.

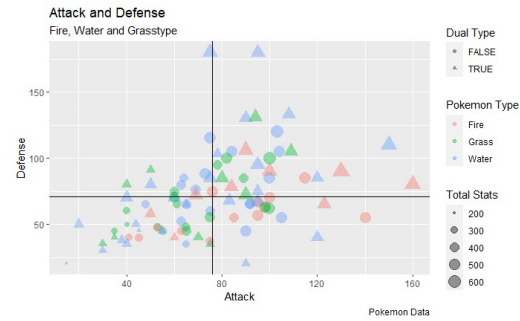


Abbildung 37 sechs Dimensionen

4.6 Facetten

Facetten sind hilfreich, um mehrere unter Plots zu generieren, um somit mehr Informationen aus der Visualisierung zu gewinnen. Die Facettierung ist hilfreich, wenn man eine kategorische Variabel oder eine diskrete quantitative Variabel als weitere Information zu den gegebenen x- und y-Achsen hinzufügen will. In Abbildung 38 erkennt man, aufgrund der Teilung einer kategorische Variabel, ob das Pokémon legendär ist oder nicht. Bei Abbildung 39 wird noch eine weitere diskrete quantitative Variabel hinzugefügt, der die Pokémon in die verschiedenen Generationen unterteilt. Schlussendlich kann man aufgrund dessen folgende Informationen gewinnen: Angriff und Verteidigung der Pokémon der Typen Feuer, Wasser und Gras. Die Form zeigt mir an, ob das Pokémon einen zweiten Typ besitzt und aufgrund der größe der Formen kann ich schliessen, wie hoch der Totale Stati ist. Zusätzlich kann ich sehen, ob es sich um ein Legendäres Pokémon handelt und aus welcher Generation diese stammt und ob diese überdurchschnittlich oder unterdurchschnittliche Attack und Defense Werte aufweist.

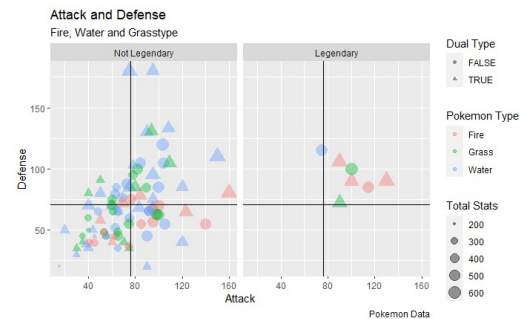


Abbildung 38 sieben Dimensionen mit `facet_wrap()`

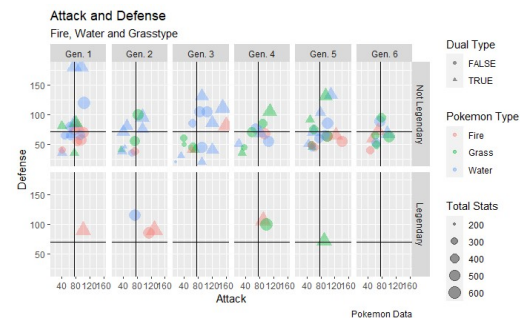


Abbildung 39 acht Dimensionen mit `facet_grid()`

4.7 Koordinaten System

Hier, in diesem Fall, stellt sich die Frage, ob die Visualisierung auf ein polares Koordinatensystem besser passt als bei einem kartesischen System. Ein polares Koordinatensystem ist, wie das kartesische Koordinatensystem, ein zweidimensionales Koordinatensystem. Jedoch befinden sich auf dem Polarkoordinatensystem Punkte in einer Ebene, der durch den Abstand von einem festgelegten Ursprung und durch den Winkel zu einer Richtung festgelegt wird („Polarkoordinaten“, 2022). Ob man das eine oder andere System nutzt, hängt von den gegebenen Daten ab, sowie die Information Übermittlung, die man mit der Visualisierung erreichen möchte. In unserem Fall reicht ein kartesisches System, um die gegebenen Informationen zu visualisieren.

5 LO5: Evaluation

Zu guter Letzt, beschäftigen wir uns in diesem Kapitel mit der Evaluation der Visualisierung. Evaluationen sind ein wichtiger Bestandteil in der Visualisierung. In der Evaluation stellt man sich die Fragen, woher man weiss, ob nun die gewünschte Visualisierung auch die entsprechenden Informationen und Effekte übermittelt und keine Nebeninformationen. Ob man die gewünschten Informationen aus der Visualisierung richtig vergleicht und entsprechend auch richtig interpretieren kann. In Bezug zur Visualisierung können wir verschiedene Evaluationen durchführen und diese entsprechend in die Visualisierung mit einfließen lassen. Eine Evaluation kann man mit verschiedener Herangehensweise evaluieren. Eine Übersicht über alle Evaluierungsmethode ist in Abbildung 40 ersichtlich.

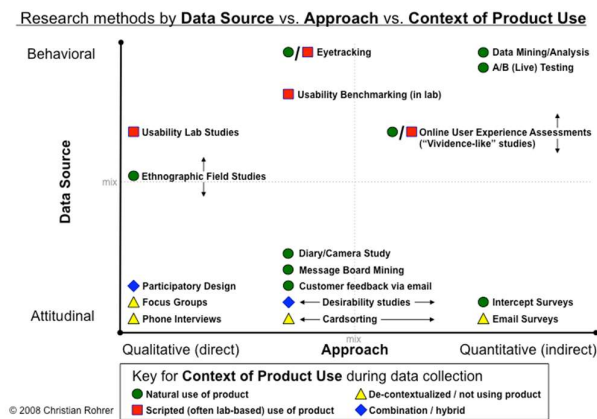


Abbildung 40 Evaluationsmethoden

5.1 Usability-Test

Bei einem Usability-test stellt man sich die Frage, ob die Visualisierung von den Menschen falsch interpretiert, werden können. Dabei ist es wichtig nicht nur eine Person zu befragen, sondern mehrere, um eine gewisse Unsicherheit der Antworten auszugleichen. In Abbildung 41 und Abbildung 42 werden die gleichen Daten in unterschiedlicher Art und Weise visualisiert und an fünf Personen gezeigt. Fünf Personen nimmt man, da man bei der ersten Person neue Informationen und Rückmeldungen erhält. Bei der zweiten Person wird es Überschneidungen geben, aber trotzdem kann neue Erkenntnisse gewonnen werden. Diese nimmt stetig mit der Grösse der Testpersonen ab (NNgroup, 2018).

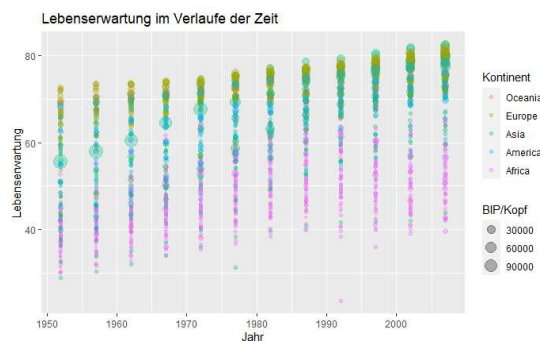


Abbildung 41

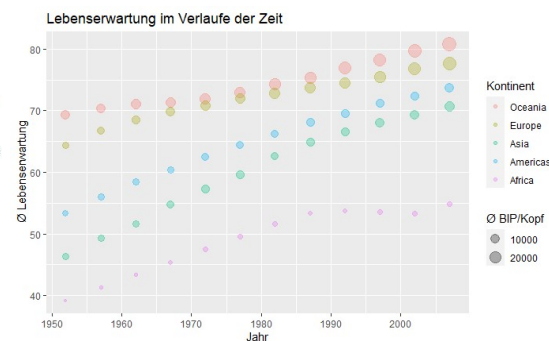


Abbildung 42

Die folgenden Fragen wurden den Probanden gestellt:

1	Beschreibe mir die Visualisierung, die du siehst.
2	Welcher Kontinent hat über den Zeitraum 1950 bis 2010 die höchste Lebenserwartung?
3	Welcher Kontinent hat über den Zeitraum 1950 bis 2010 die niedrigste Lebenserwartung?
4	Hatte der Kontinent Afrika im Jahre 1982 einen höheren BIP/Kopf als Asien?
5	Kannst du einen Zusammenhang zwischen BIP/Kopf und Lebenserwartung sehen?
6	Welche Visualisierung findest du Optisch schöner?
7	Warum findest du diese Visualisierung besser?

5.2 Durchführung der Umfrage

Für die Umfragen wurden fünf unterschiedliche Personen befragt, mit unterschiedlichen Lehrausbildungen und Bildungsniveau. Das Alter der Probanden liegt zwischen 20 bis 25 Jahren. Zwei Personen waren männlich und drei Personen weiblich. Die Befragung fand einzeln unter vier Augen in einem ruhigen, ungestörten Raum statt und dauerten im Durchschnitt zehn Minuten. Zuerst wurde Abbildung 41 präsentiert und die Fragen 1 bis 5, der Reihe nachgestellt. Anschliessend wurde Abbildung 42 präsentiert und die Fragen analog zu Abbildung 41 wiederholt. Für die ersten fünf Fragen, jeweils für beide Abbildungen, wurde ein Timer gestellt. Der Proband hatte nun beide Visualisierungen vor sich und die Fragen 6 und 7 wurden gestellt. Die Antworten wurden auf einem Notizblock festgehalten, für die spätere Auswertung.

5.3 Auswertung der Umfrage

Von den fünf ausgewählten Personen konnten alle die Frage beantworten. Da fünf Personen eine kleine Personengruppe umfasst, ist diese nicht stark reproduzierbar, jedoch lässt sich aufgrund dessen schon einmal eine kleine Auswertung machen. Um Hypothesen und Behauptungen belegen oder widerlegen zu können, müsste man eine grössere Personengruppe befragen.

Die erste Frage konnten allen Probanden sehr gut, unabhängig von den Abbildungen, beantworten. Die meisten haben den Titel der Visualisierung sowie die x- und y-Achsen Beschriftungen beschrieben. Aufgrund der Legenden konnten alle Probanden die Kontinente aufgrund der Farbe zuzuweisen sowie feststellen, dass die Grösse der Punkte oder Linien mit dem BIP/Kopf zusammenhängt. Auch wurde erwähnt, dass sich die Lebenserwartung im Verlaufe der Zeit stieg. Interessant ist, dass Abbildung 41 als Liniendiagramm interpretiert wurde statt als Punktediagramm.

Die zweite Frage konnte in der Abbildung 41 nur schwer beantwortet werden, da ab dem Jahr 1982, die Punkte von Asien über den von Europa liegt. Die meisten gaben hier als Antwort Europa an. Anders sah es dann bei Abbildung 42 aus, da war die klare Antwort Ozeanien.

Die dritte Frage wurden in beiden Abbildungen mit Afrika beantwortet. Die meisten gaben auch an, dass diese mit der jetzigen wirtschaftlichen Situation auch plausibel ist, unabhängig von der Visualisierung. Bei dieser Frage hat nicht die Visualisierung die Antwort beeinflusst, sondern die Vorbildung der jeweiligen Person.

Die vierte Frage wurde in Abbildung 41 verzögert beantwortet und dauerte im Schnitt länger als bei Abbildung 42. Dies hat auch damit zu tun, dass die Punkte als verschwommene Linien betrachtet wurde und sich in Abbildung 41 überlappen. Drei Probanden waren unsicher, haben aber die Frage nach einiger Zeit korrekt beantwortet.

Die fünfte Frage wurde bei Abbildung 41 anders beantwortet als bei Abbildung 42. Bei Abbildung 41 gaben die meisten an, dass je dünner die Linien unten im Diagramm sind, desto kleiner das BIP/Kopf des jeweiligen Kontinentes ist. Bei Abbildung 42 wurde klar, dass die Grösse der Punkte mit dem BIP/Kopf zusammenhängt. Im Schnitt dauerten die ersten fünf Fragen mit der Abbildung 41 länger als bei Abbildung 42.

Bei der sechsten Frage, waren alle Probanden der klaren Meinung, dass Abbildung 42 rein optisch betrachtet, schöner ist. Eine Person gab als Bemerkung an, dass der Inhalt der Legende auf Englisch beschrieben ist und der Rest auf Deutsch.

Bei der siebten Frage gaben die meisten an, dass das Interpretieren der Visualisierung in Abbildung 41, bzw. die Beantwortung der Fragen und das Herauslesen von Informationen viel schwieriger ist, da sich die Farben und Grössen untereinander vermischen und somit diese nicht klar auf einen Blick ersichtlich ist.

6 Anhang Abbildungen

Abbildung: 1 Scheinkonturen



Abbildung: 2 Streudiagramm

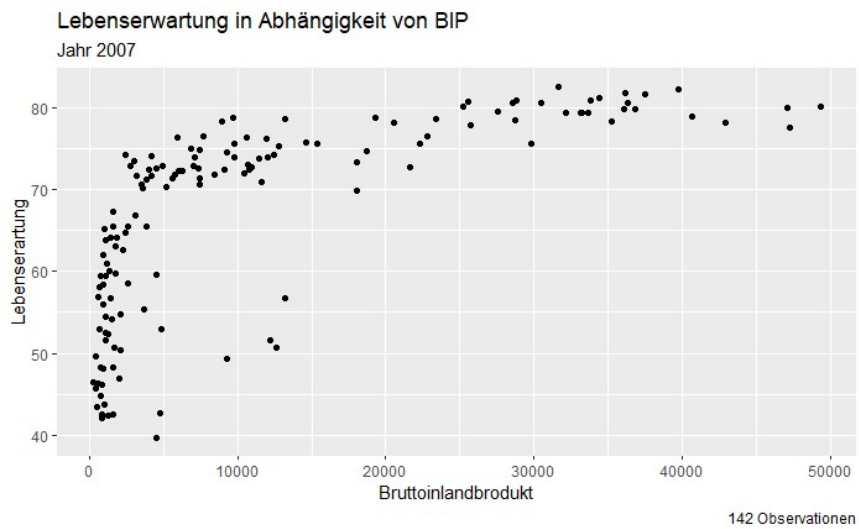


Abbildung: 3 Streudiagramm mit Kategorische Variabel

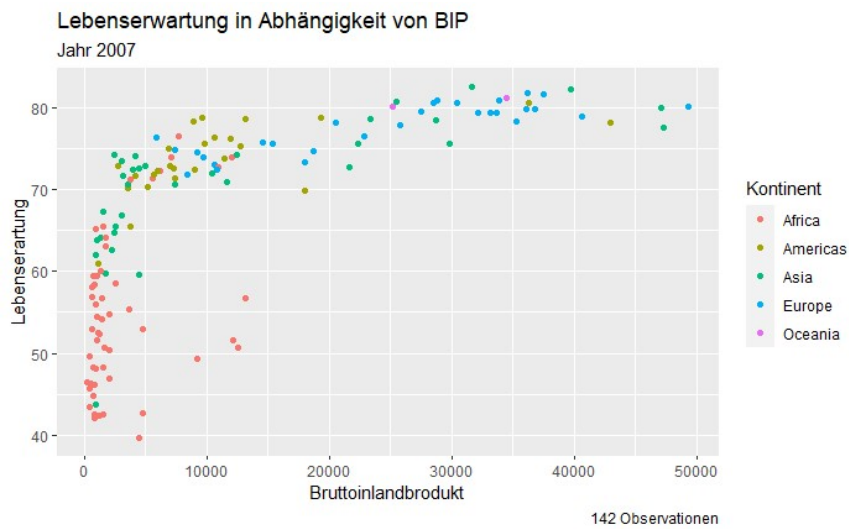


Abbildung: 4 Blasendiagramm

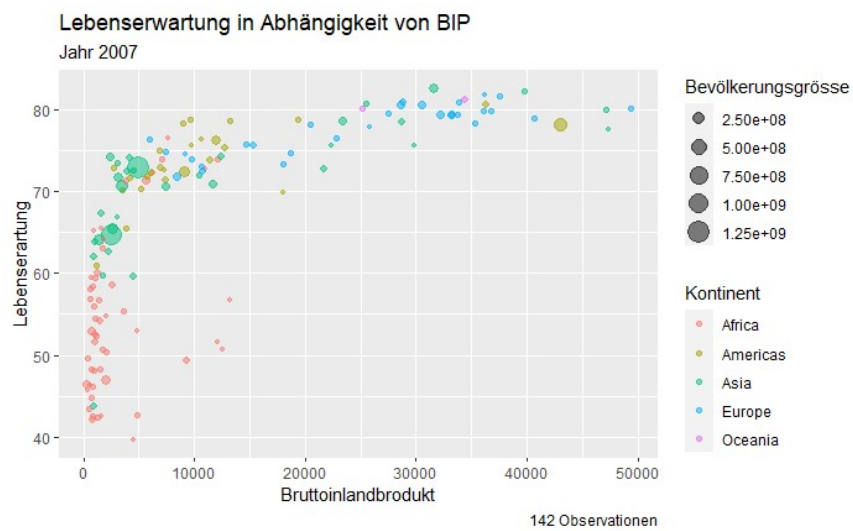


Abbildung: 5 Liniendiagramm

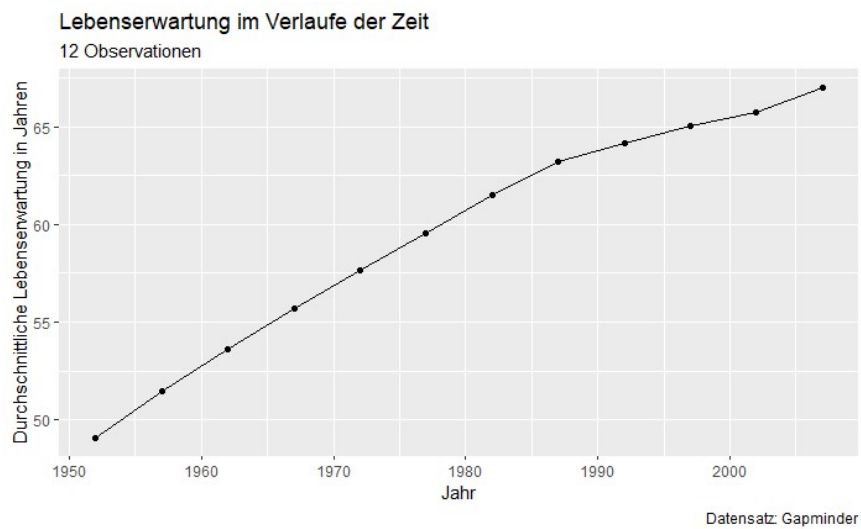


Abbildung: 6 Liniendiagramm mit Kategorische Variabel

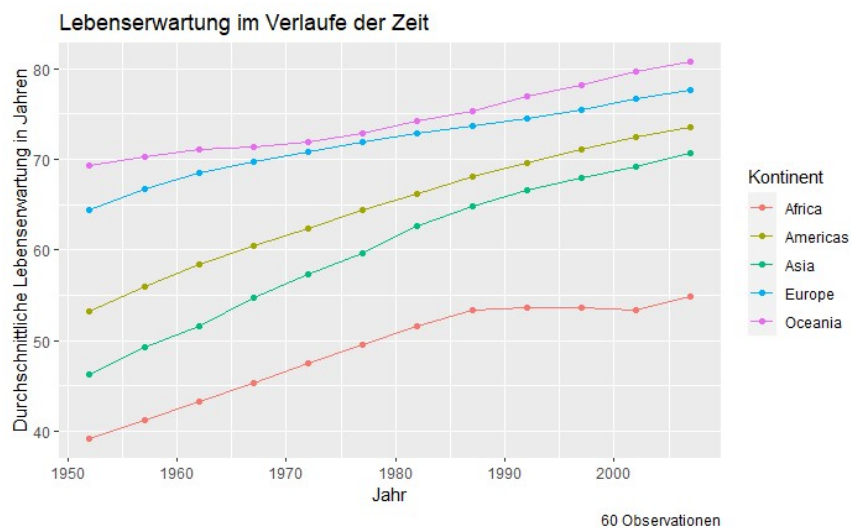


Abbildung: 7 Anzahl bins = 3

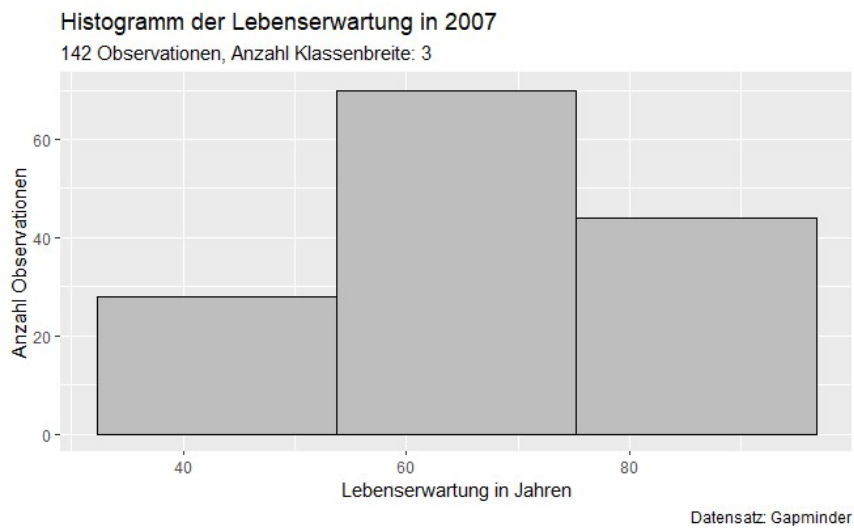


Abbildung: 8 Anzahl bins = 142

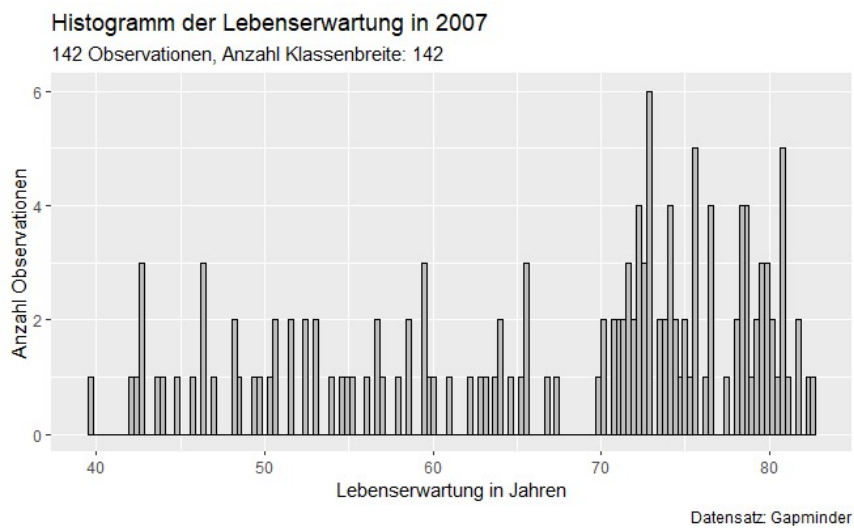


Abbildung: 9 Anzahl bins = 30

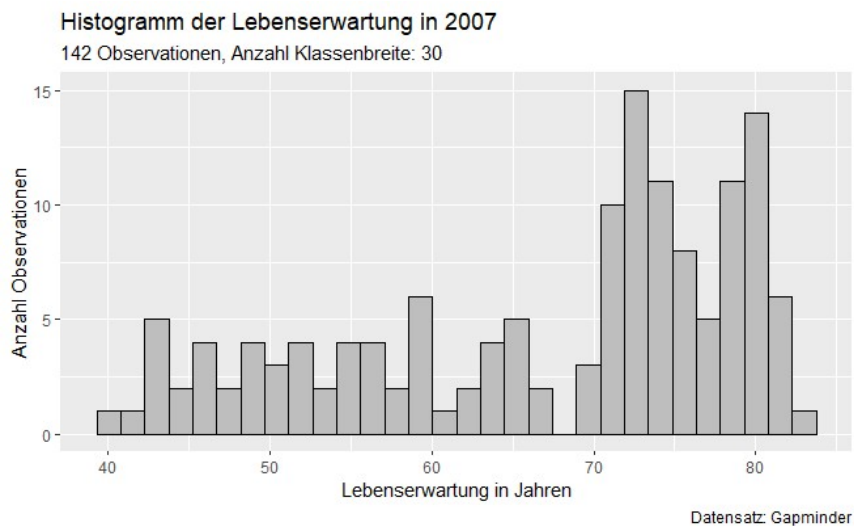


Abbildung: 10 Aufgestapelt

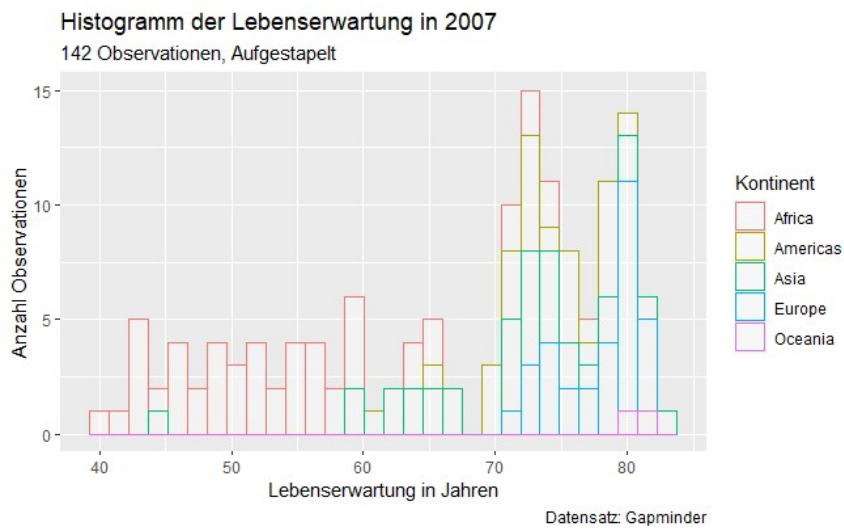


Abbildung: 11 Überlappung

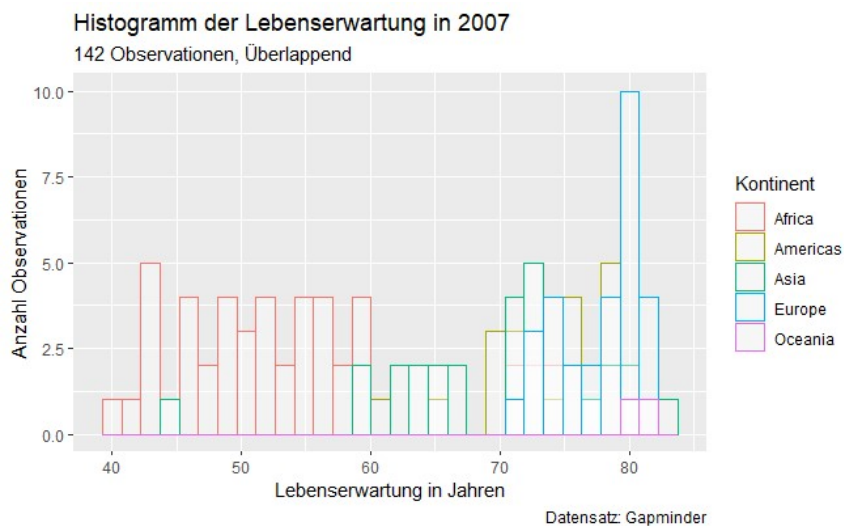


Abbildung: 12 Nebeneinander

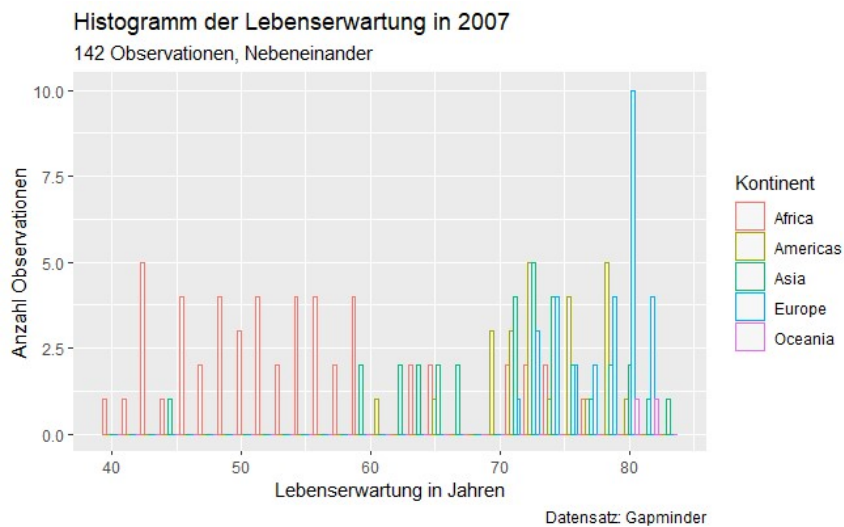


Abbildung: 13 Grafik vor Anpassung

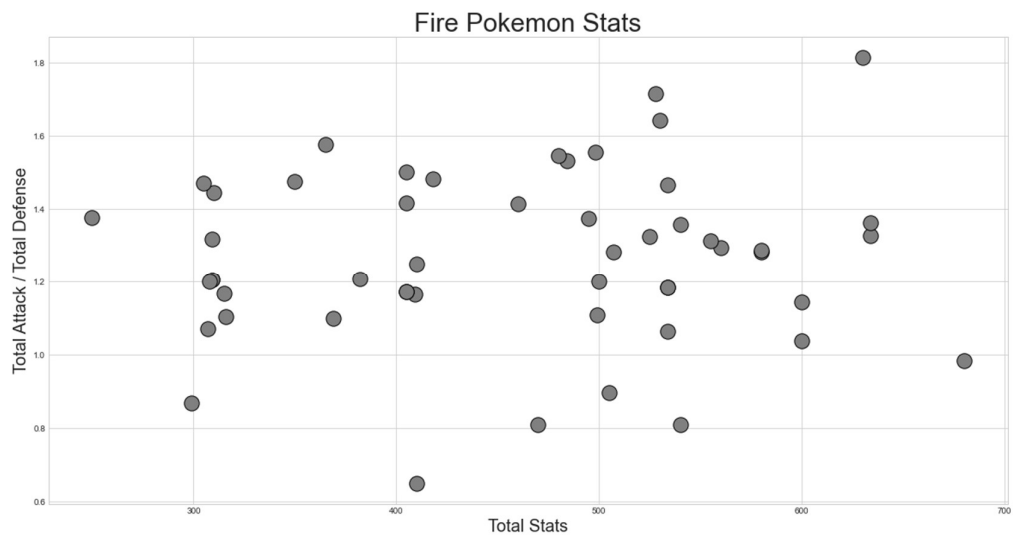


Abbildung: 14 Grösse

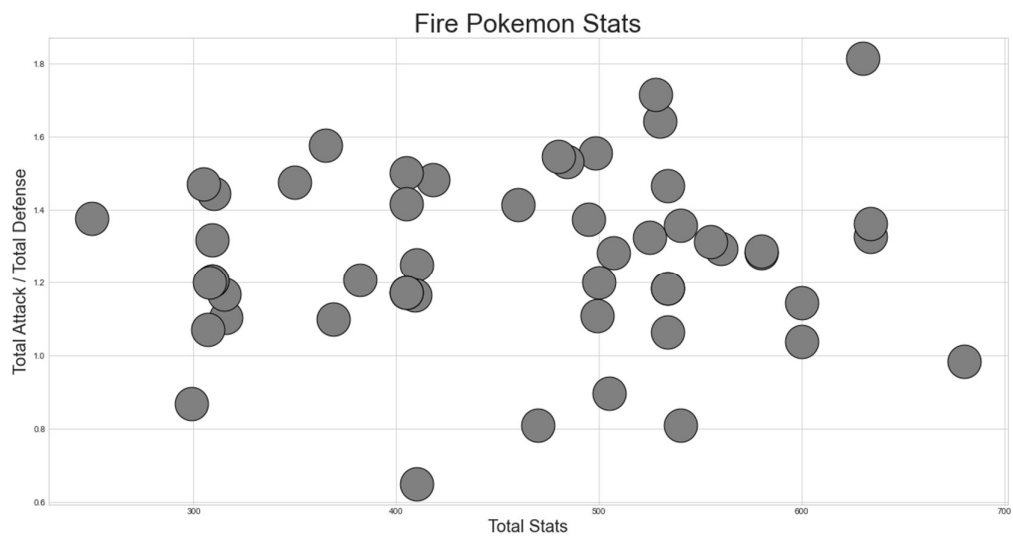


Abbildung: 15 Skalen Logarithmieren

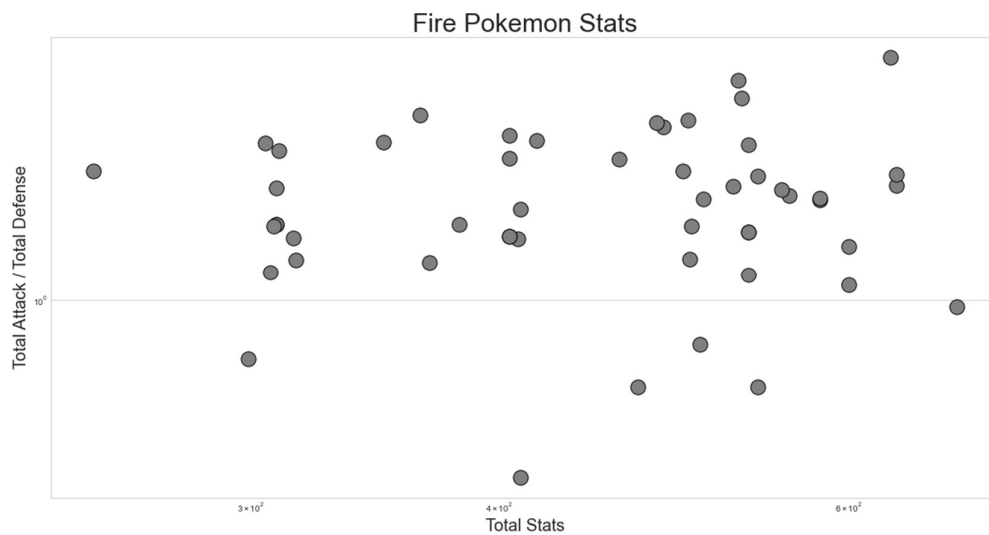


Abbildung: 16 x- und y-Achse vertauschen

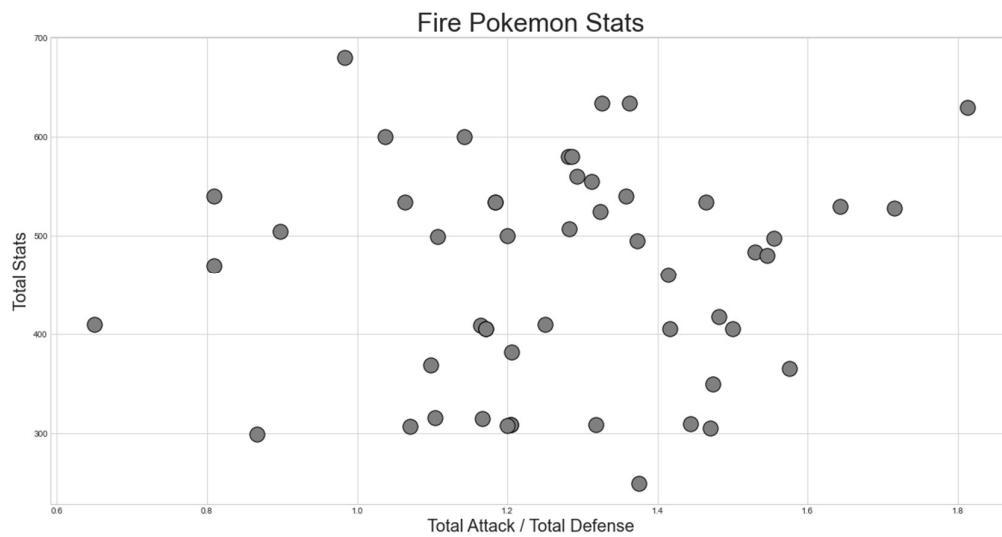


Abbildung: 17 Quadratische Formen

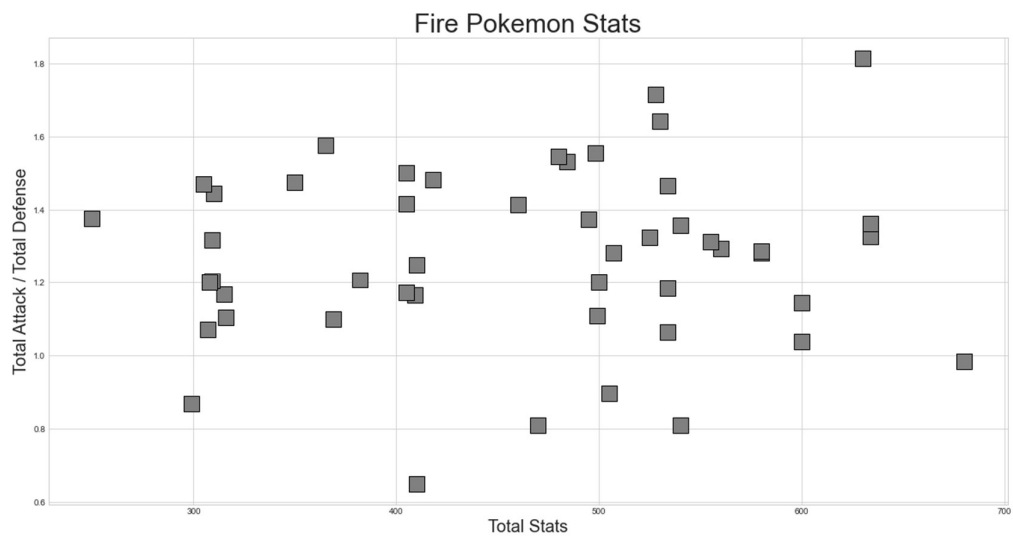


Abbildung: 18 Kreuz Formen

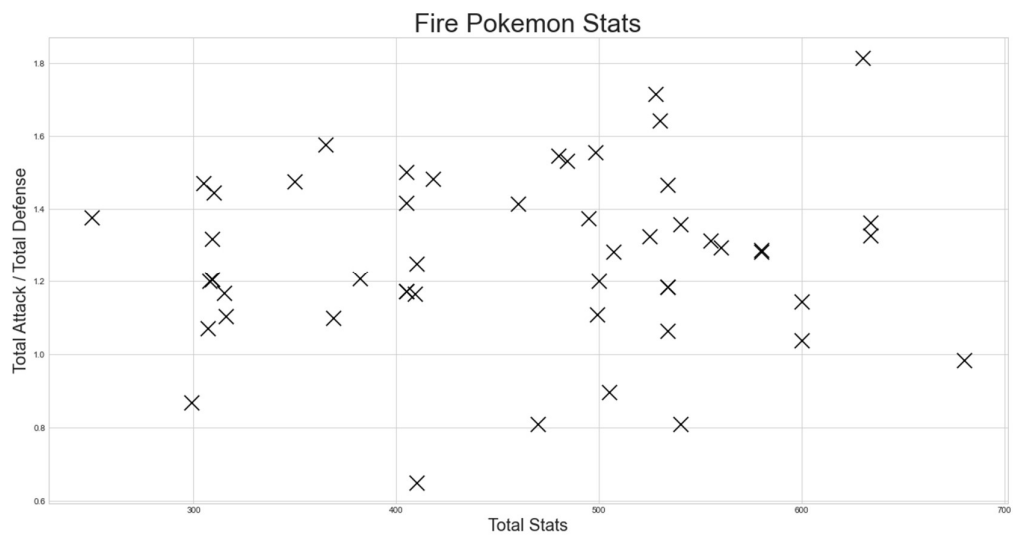


Abbildung: 19 Transparenz

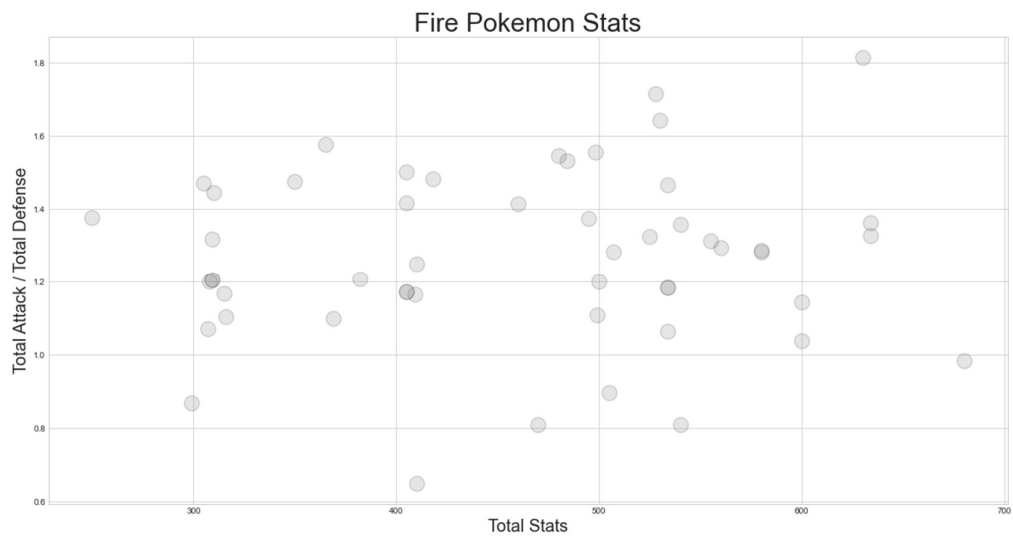


Abbildung: 20 Transparenz mit Blau

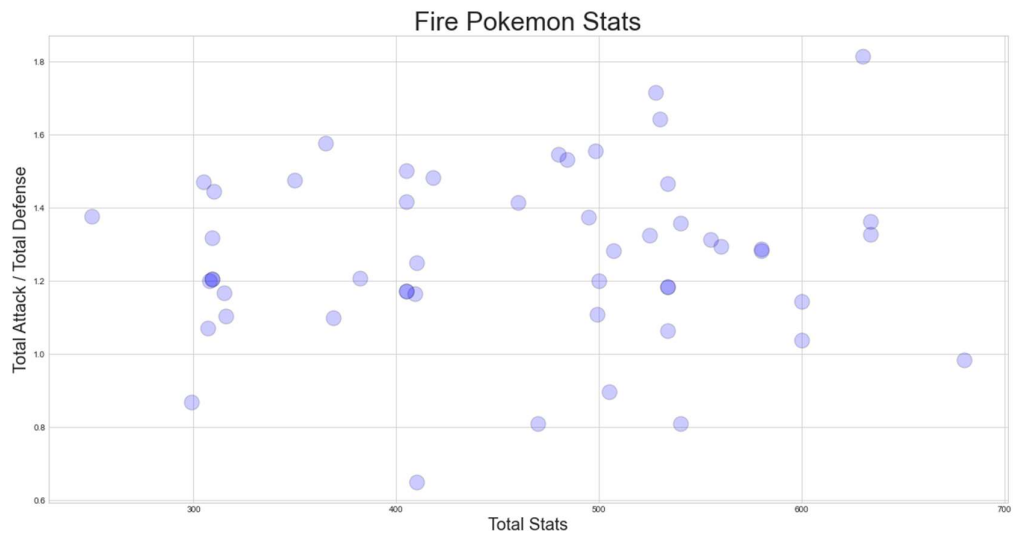


Abbildung: 21 eingefärbt mit Rot

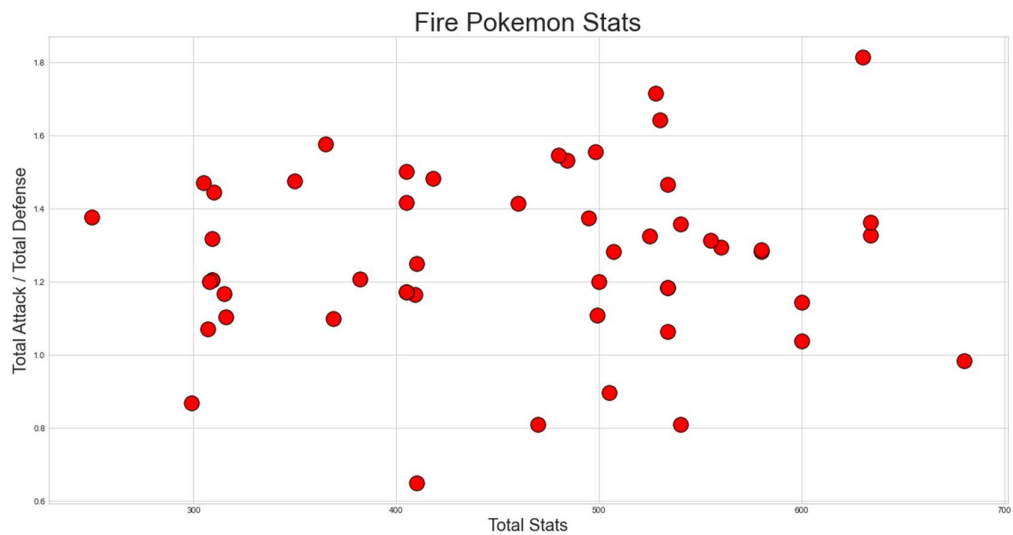


Abbildung: 22 eingefärbt aufgrund kategorische Variabel

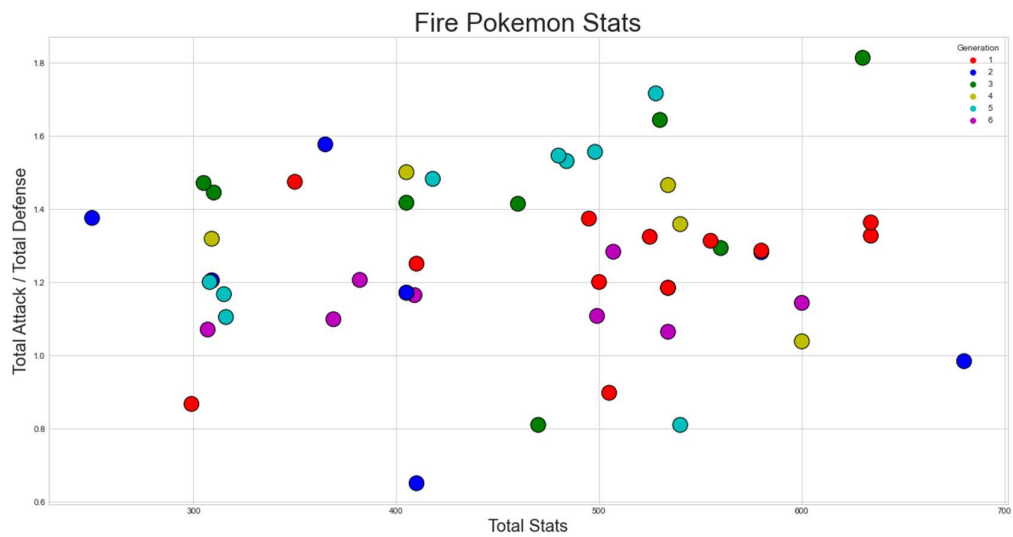


Abbildung: 23 ø Punktdiagramm

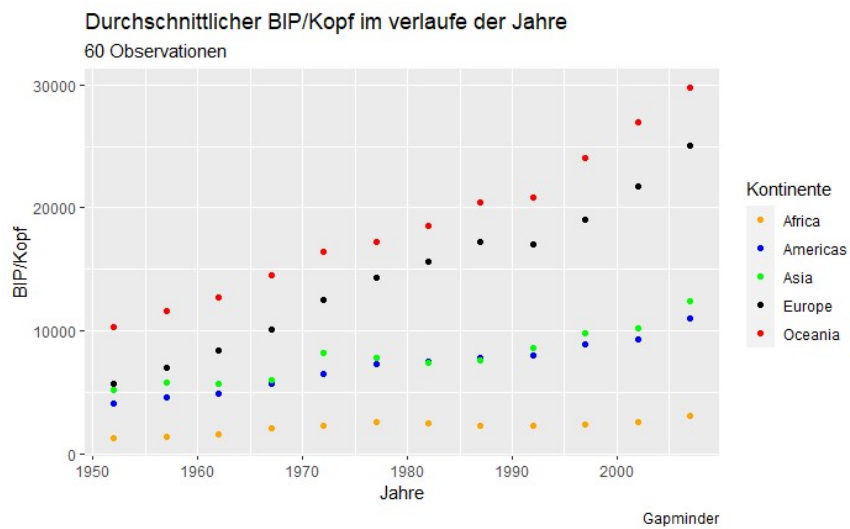


Abbildung: 24 ø Liniendiagramm

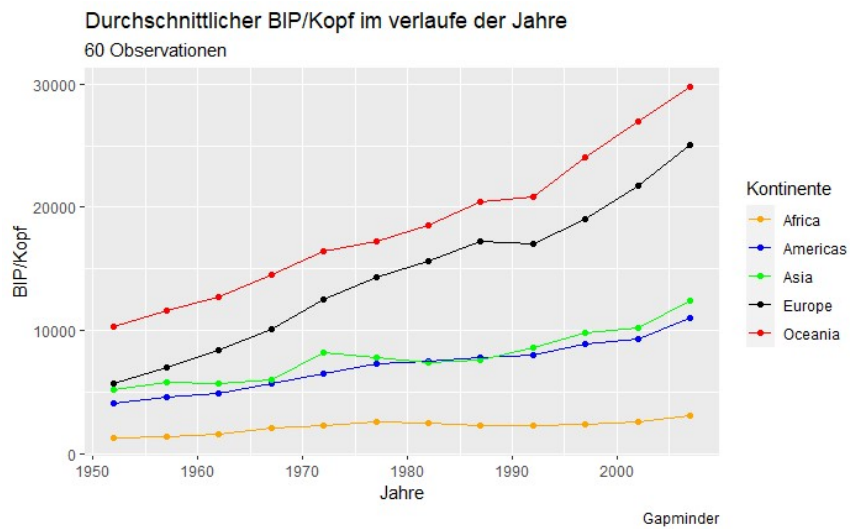


Abbildung: 25 Median Punktdiagramm

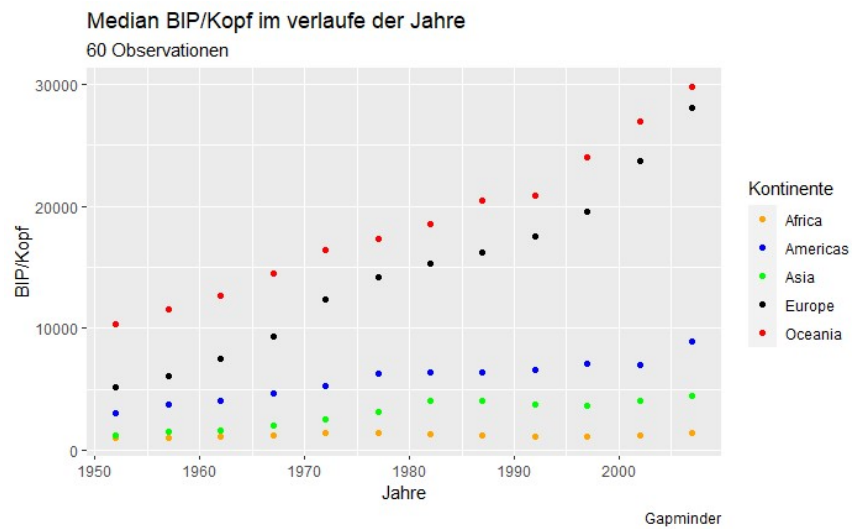


Abbildung: 26 Median Liniendiagramm

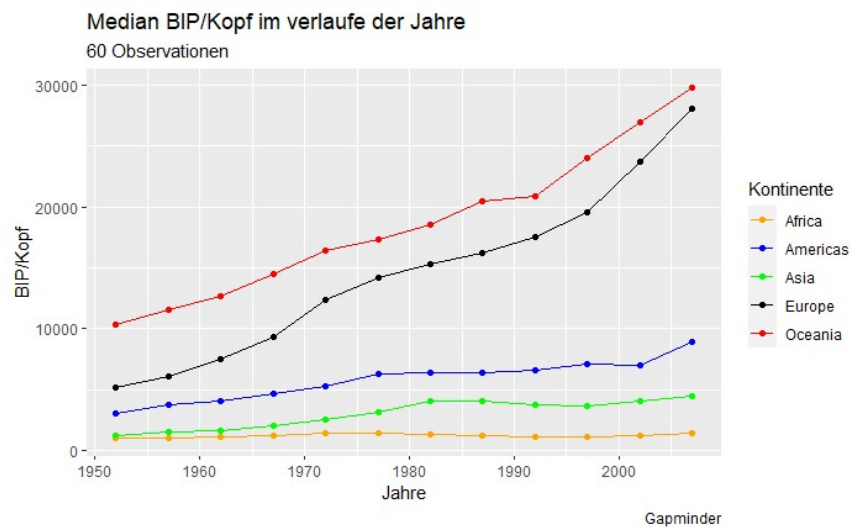


Abbildung: 27 Skalierung der y-Achse (0 – 30000)

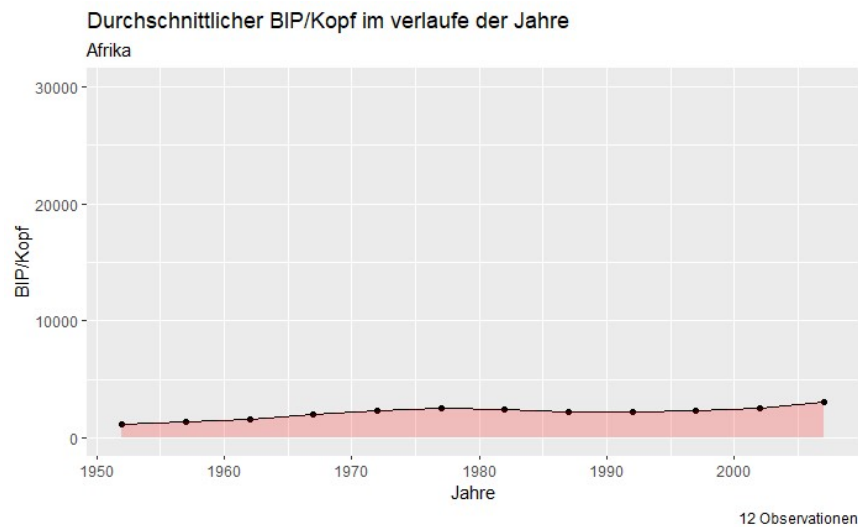


Abbildung: 28 Skalierung der y-Achse (0 – 3000)

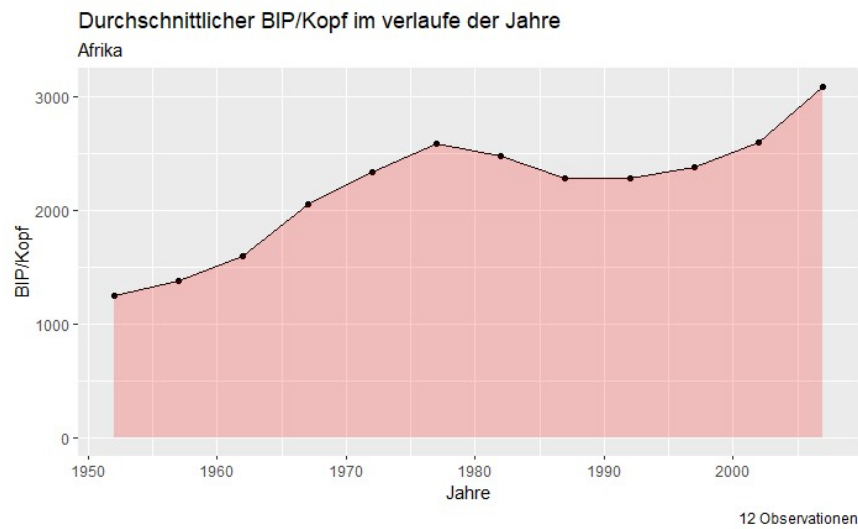


Abbildung: 29 Legende ohne Reihenfolge

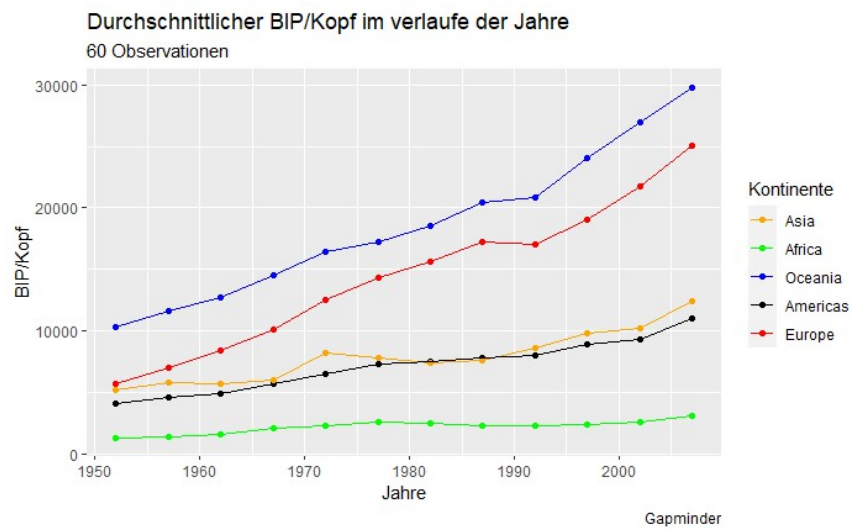


Abbildung: 30 Legende ohne Reihenfolge

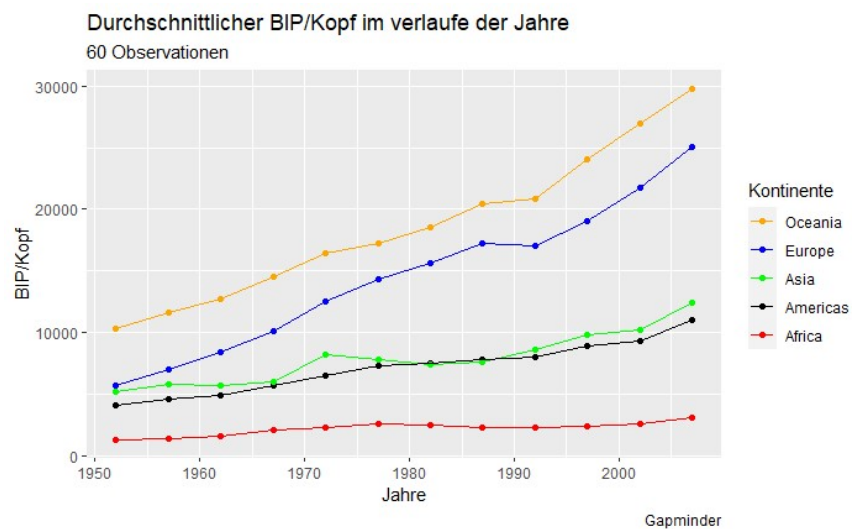


Abbildung: 31 „Hässliche“ Tabelle

country	continent	year	lifeExp	pop	gdpPercap
Japan	Asia	2007	82.603	127467972	31656.0681
Hong Kong	Asia	2007	82.208	6980412	39724.9787
Iceland	Europe	2007	81.757	301931	36180.7892
Switzerland	Europe	2007	81.701	7554661	37506.4191
Australia	Oceania	2007	81.235	20434176	34435.3674

Abbildung: 32 „Schöne“ Tabelle

country	continent	year	lifeExp	pop	gdpPercap
Japan	Asia	2007	82.603	127467972	31656.0681
Hong Kong	Asia	2007	82.208	6980412	39724.9787
Iceland	Europe	2007	81.757	301931	36180.7892
Switzerland	Europe	2007	81.701	7554661	37506.4191
Australia	Oceania	2007	81.235	20434176	34435.3674

Abbildung: 33 Zwei Dimensionen

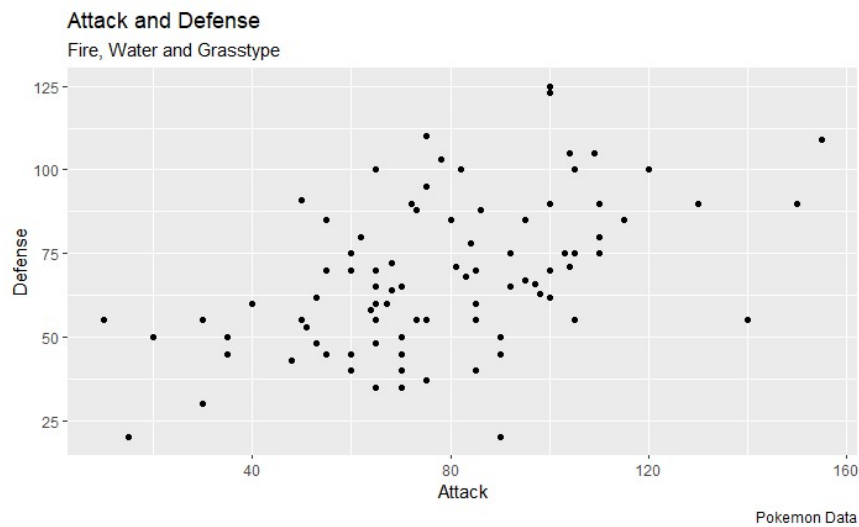


Abbildung: 34 Drei Dimensionen

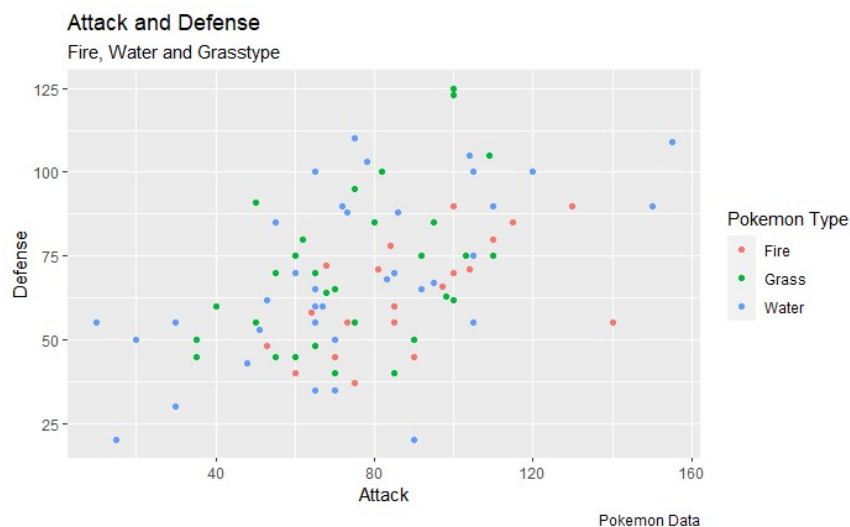


Abbildung: 35 Vier Dimensionen

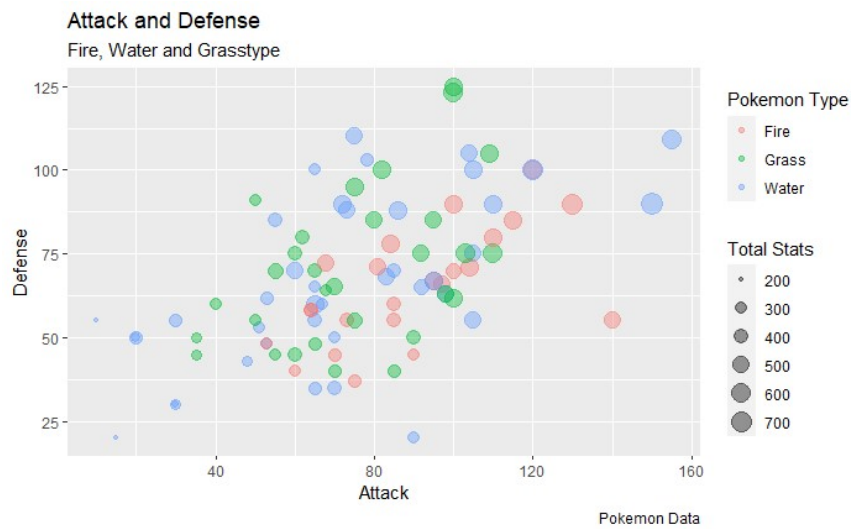


Abbildung: 36 fünf Dimensionen

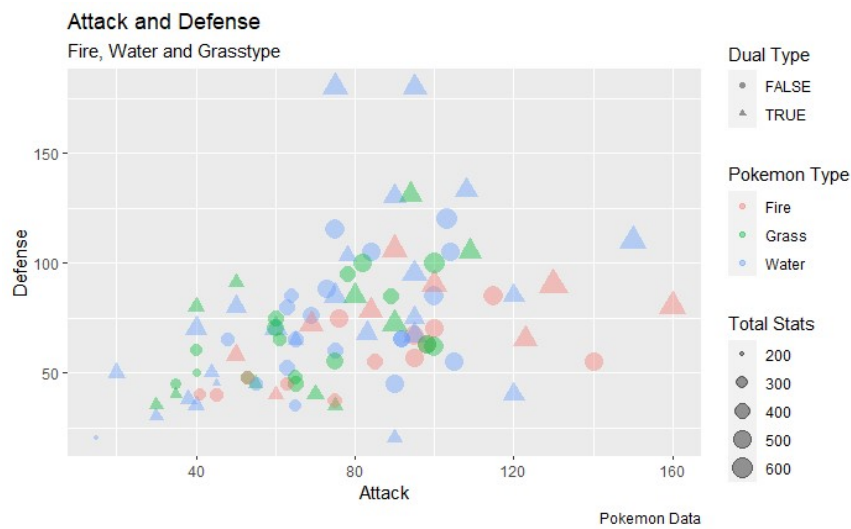


Abbildung: 37 sechs Dimensionen

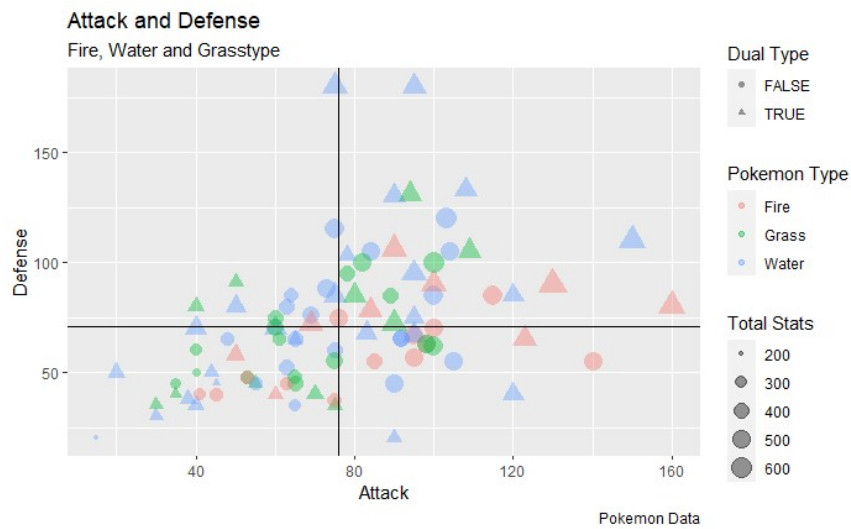


Abbildung: 38 sieben Dimensionen mit `facet_wrap()`

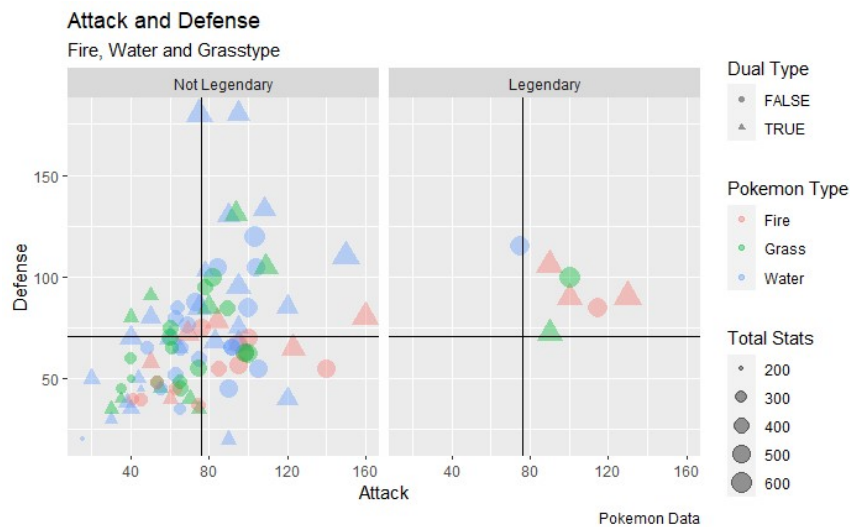


Abbildung: 39 acht Dimensionen mit `facet_grid()`

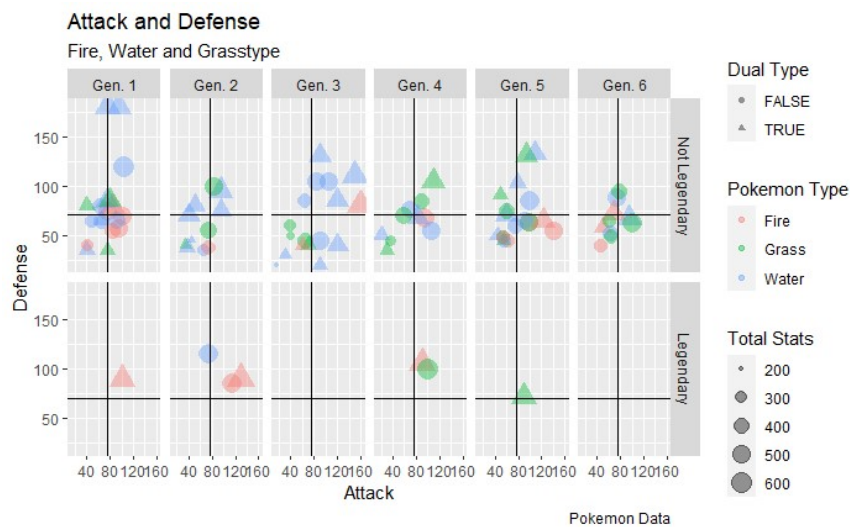


Abbildung: 40 Evaluationsmethoden

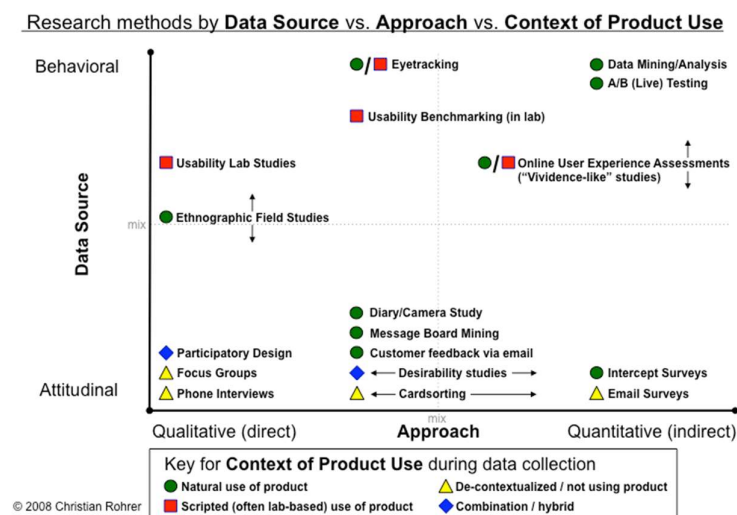


Abbildung: 41

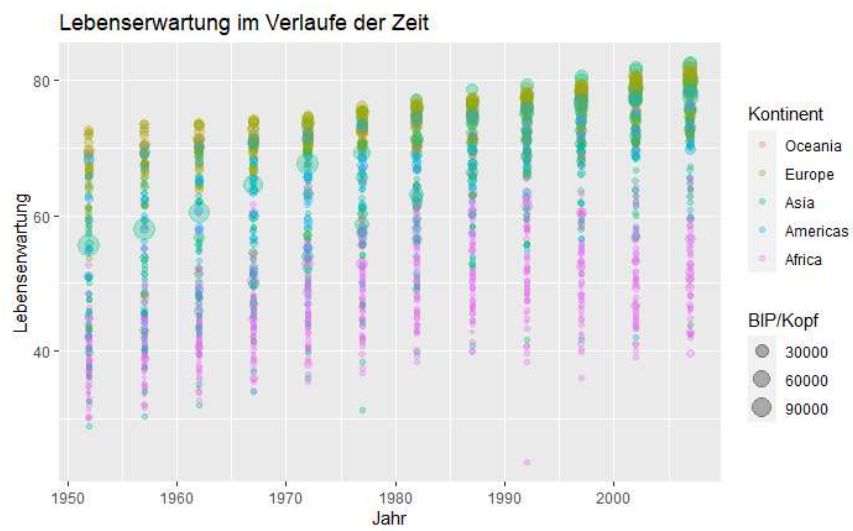
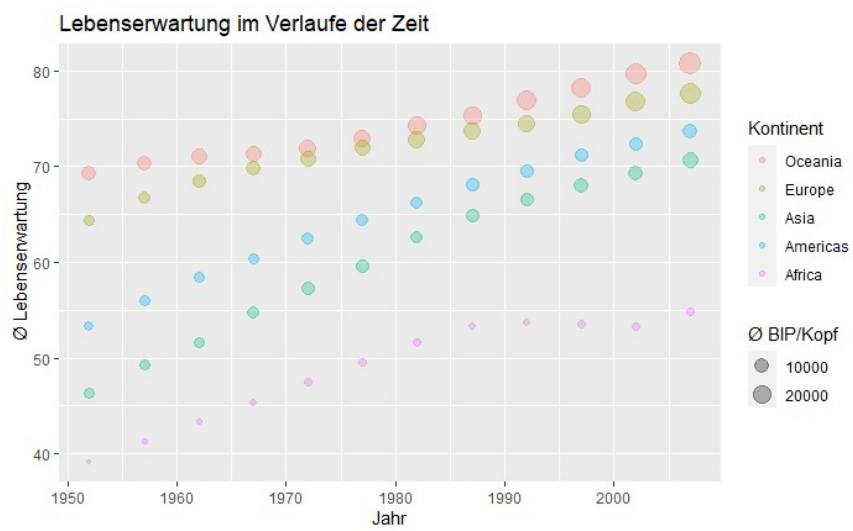


Abbildung: 42



7 Quellenverzeichnis

- Anton, L. B. & T. (2019). *Kapitel 9 Visualisierung | R für Psychos*. <https://r-intro.tadaa-data.de/book/>
- Bedeutung der Farben ["Farbsymbolik und Farben in der Psychologie"]. (2020, Januar 5). Designer in Action. <https://www.designerinaction.de/design-wissen/bedeutung-farben/>
- Broese, S. A., Antwerpes, F., Fink, B., & Ostendorf, N. (2020, Januar 24). *Streudiagramm* [Informationswebsite]. DocCheck Flexikon. <https://flexikon.doccheck.com/de/Streudiagramm>
- Bryan, J. (Jenny), & MacDonald, A. (2015). *Gapminder: Turned On Integration With Zenodo*. Zenodo. <https://doi.org/10.5281/ZENODO.594018>
- Clemens, P. (2016, Oktober 24). *Über die Subjektivität unserer Wahrnehmung und der Wirklichkeit*. <https://uni.de/redaktion/subjektive-wahrnehmung>
- Cramer und Kamps—2020—Grundlagen der Wahrscheinlichkeitsrechnung und Sta.pdf. (o. J.). Abgerufen 20. März 2022, von <https://link.springer.com/content/pdf/10.1007%2F978-3-662-60552-3.pdf>
- Ferreira, C. R. (2021, April 2). *Two Simple Steps to Create Colorblind Friendly Data Visualizations*. Medium. <https://towardsdatascience.com/two-simple-steps-to-create-colorblind-friendly-data-visualizations-2ed781a167ec>
- Handbuch zur Datenvisualisierung: Definition, Beispiele und Lernressourcen. (2022, März 14). [Lernplattform]. Tableau. <https://www.tableau.com/de-de/learn/articles/data-visualization>
- Jugel, U., Jerzak, Z., Hackenbroich, G., & Markl, V. (2014). M4: A visualization-oriented time series data aggregation. *Proceedings of the VLDB Endowment*, 7(10), 797–808. <https://doi.org/10.14778/2732951.2732953>
- Kurven- und Punktdiagramme einfach erklärt mit unserer Hilfe. (o. J.). Abgerufen 14. März 2022, von <https://learnattack.de/schuelerlexikon/mathematik/kurvendiagramme>
- NNgroup. (2018, Oktober 26). *Usability Testing w. 5 Users: Design Process (video 1 of 3)*. <https://www.youtube.com/watch?v=RhgUirqki50>
- Perruchoud, D. (2022). *Datenvisualisierung mit R*. 44.

Polarkoordinaten. (2022). In *Wikipedia*.

<https://de.wikipedia.org/w/index.php?title=Polarkoordinaten&oldid=219706397>

Rounak, B. (2017). *The Complete Pokemon Dataset*. <https://kaggle.com/rounakbanik/pokemon>

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>

Scheinkonturen. (2022). In *Wikipedia*.

<https://de.wikipedia.org/w/index.php?title=Scheinkonturen&oldid=219241882>

Streu- und Blasendiagramme—Hilfe für Data Studio. (o. J.). Abgerufen 20. März 2022, von

<https://support.google.com/datastudio/answer/7207785?hl=de#zippy=%2Cthemen-in-diesem-artikel>

Ware, C. (2013). *Information Visualization: Perception for Design*. Elsevier.

Wie findet man das richtige Diagramm für seine Daten? (o. J.). Infogram. Abgerufen 14. März 2022, von <https://infogram.com/de/de/seite/waehlen-sie-das-richtige-diagramm-fuer-ihre-datenvisualisierung>

Wilke, C. O. (2019). *Fundamentals of Data Visualization*. <https://clauswilke.com/dataviz/>

Yi, M. (2021). *A Complete Guide to Scatter Plots*. Chartio. <https://chartio.com/learn/charts/what-is-a-scatter-plot/>

8 Abbildungsverzeichnis

ABBILDUNG: 1 SCHEINKONTUREN	13
ABBILDUNG: 2 STREUDIAGRAMM	13
ABBILDUNG: 3 STREUDIAGRAMM MIT KATEGORISCHE VARIABLE	13
ABBILDUNG: 4 BLASENDIAGRAMM	14
ABBILDUNG: 5 LINIENDIAGRAMM	14
ABBILDUNG: 6 LINIENDIAGRAMM MIT KATEGORISCHE VARIABLE	14
ABBILDUNG: 7 ANZAHL BINS = 3	15
ABBILDUNG: 8 ANZAHL BINS = 142	15
ABBILDUNG: 9 ANZAHL BINS = 30	15
ABBILDUNG: 10 AUFGESTAPELT	16
ABBILDUNG: 11 ÜBERLAPPUNG	16
ABBILDUNG: 12 NEBENEINANDER	16
ABBILDUNG: 13 GRAFIK VOR ANPASSUNG	17
ABBILDUNG: 14 GRÖSSE	17
ABBILDUNG: 15 SKALEN LOGARITHMIEREN	17
ABBILDUNG: 16 X- UND Y-ACHSE VERTAUSCHEN	18
ABBILDUNG: 17 QUADRATISCHE FORMEN	18
ABBILDUNG: 18 KREUZ FORMEN	18
ABBILDUNG: 19 TRANSPARENZ	19
ABBILDUNG: 20 TRANSPARENZ MIT BLAU	19
ABBILDUNG: 21 EINGEFÄRBT MIT ROT	19
ABBILDUNG: 22 EINGEFÄRBT AUFGRUND KATEGORISCHE VARIABLE	20
ABBILDUNG: 23 Ø PUNKTDIAGRAMM	20
ABBILDUNG: 24 Ø LINIENDIAGRAMM	20
ABBILDUNG: 25 MEDIAN PUNKTDIAGRAMM	21
ABBILDUNG: 26 MEDIAN LINIENDIAGRAMM	21
ABBILDUNG: 27 SKALIERUNG DER Y-ACHSE (0 – 30000)	21
ABBILDUNG: 28 SKALIERUNG DER Y-ACHSE (0 – 3000)	22
ABBILDUNG: 29 LEGENDE OHNE REIHENFOLGE	22
ABBILDUNG: 30 LEGENDE OHNE REIHENFOLGE	22
ABBILDUNG: 31 „HÄSSLICHE“ TABELLE	23
ABBILDUNG: 32 „SCHÖNE“ TABELLE	23
ABBILDUNG: 33 ZWEI DIMENSIONEN	23
ABBILDUNG: 34 DREI DIMENSIONEN	23
ABBILDUNG: 35 VIER DIMENSIONEN	24
ABBILDUNG: 36 FÜNF DIMENSIONEN	24
ABBILDUNG: 37 SECHS DIMENSIONEN	24
ABBILDUNG: 38 SIEBEN DIMENSIONEN MIT FACET_WRAP()	25
ABBILDUNG: 39 ACHT DIMENSIONEN MIT FACET_GRID()	25
ABBILDUNG: 40 EVALUATIONSMETHODEN	25
ABBILDUNG: 41	26
ABBILDUNG: 42	26

9 Ehrlichkeitserklärung

Hiermit erkläre ich, das vorliegende Dokument «Grundlagen Datenvisualisierung» selbständig und unter Benutzung der angegebenen Quellen verfasst zu haben. Die wörtlich oder inhaltlich aus den aufgeführten Quellen entnommenen Stellen sind in der Arbeit als Zitat bzw. Paraphrase kenntlich gemacht.

Das Dokument dient als Leistungsnachweis für das Modul «Grundlagen Datenvisualisierung (gdv)».

Windisch, 24. Mai .2022

Name: Si Ben Tran

Unterschrift:

A handwritten signature in black ink, appearing to be 'SBT' or similar, written in a cursive style.