**Grundkompetenz Datenvisualisierung**

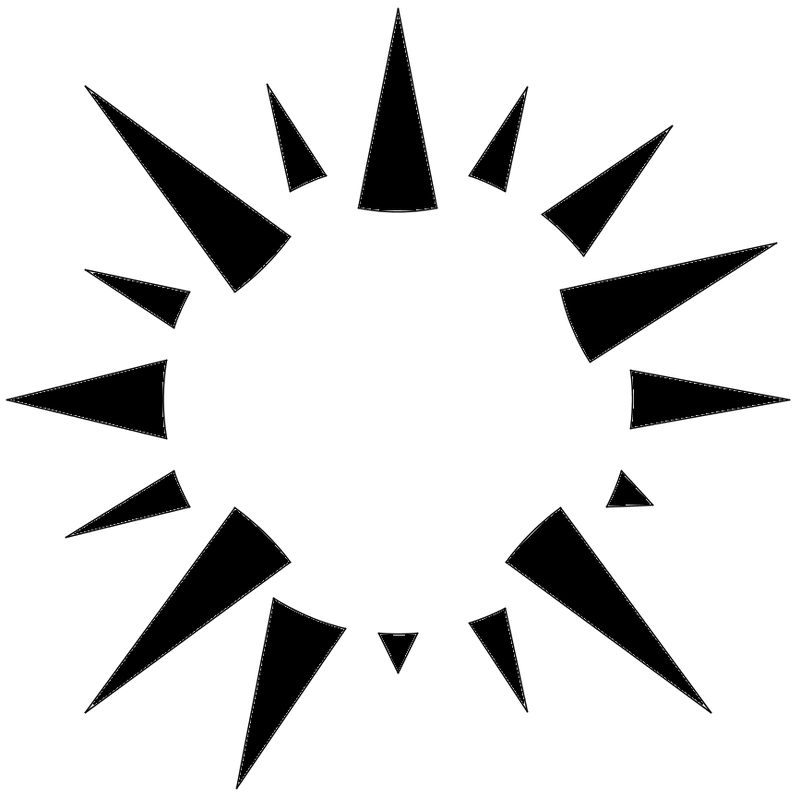


Abbildung 1: https://de.wikipedia.org/wiki/Scheinkonturen#/media/Datei:Scheinkonturen\_Kugel.png

Student Si Ben Tran

Expertin Prof. Dr. Arzu Çöltekin

GitHub https://github.com/7ben18/grundlagen\_datenvisualisierung

Fachhochschule Nordwestschweiz, Hochschule für Technik

Windisch, 17. Juni 2022

Inhaltsverzeichnis

[1 LO1: Visualization basics, chart types 3](#_Toc97113579)

[1.1 Diagramm Beschriftung 3](#_Toc97113580)

[1.2 Streudiagramm - Liniendiagramm 3](#_Toc97113581)

[1.3 Histogramme 4](#_Toc97113582)

[2 LO2: Visual Perception 5](#_Toc97113583)

[3 LO3: Design Principles vs. Data 5](#_Toc97113584)

[4 LO4: Grammar of Graphics Tools 5](#_Toc97113585)

[5 LO5: Evaluation 5](#_Toc97113586)

[6 Anhang Abbildungen 5](#_Toc97113587)

# LO1: Visualization basics, chart types

Visualisierungen können sehr hilfreich sein, um in kurzer Zeit einen Überblick über den Datensatz zu erhalten. Dabei ist es wichtig, die grundlegenden Visualisierungsgrundlagen zu beherrschen und entsprechend auch richtig anzuwenden und zu verstehen. In diesem Kapitel geht es darum, die grundlegenden Visualisierungsarten zu zeigen und was bei einer Visualisierung zu beachten ist, damit Informationen der Daten nicht verloren gehen. Dabei werden wir den Datensatz von 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 Bruttoinlandprodukt, unterteilt in Länder und Kontinente jeweils in einem Zeitintervall in 5 Jahren von 1952 bis 2007 beinhaltet.

## **Diagramm Beschriftung**

Erst durch eine Beschriftung, wird der Visualisierung Leben eingehaucht. Ein Diagramm, dass keine oder nur unvollständige Beschriftungen hat, ist sehr schwer zu interpretieren, da man keinen Anhaltspunkt hat welche Informationen die Visualisierung übermitteln soll (Abbildung 2).

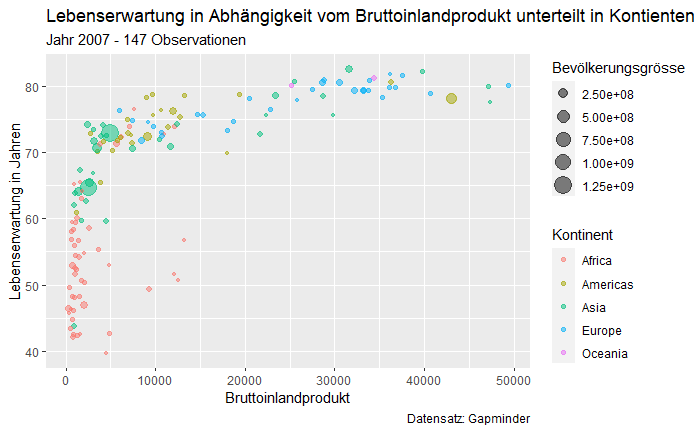
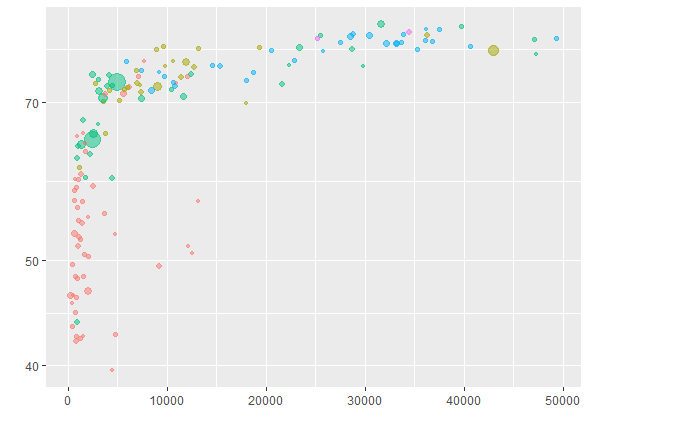


Abbildung 2 Diagramm ohne Beschriftung Abbildung 3 Diagramm mit Beschriftung

Zu einer Beschriftung essenziell wichtig ist ein Titel und die Achsenbeschriftungen. Mit diesen Informationen weiss man auf den ersten Blick, welche Thematik die Visualisierung beinhaltet. Einen Untertitel sowie Caption und Beschriftungen der Legenden sind sinnvoll, wenn viele Informationen auf einen Plot dargestellt werden, dies erleichtert das Interpretieren der Visualisierung (Abbildung 3).

## **Streudiagramm - Liniendiagramm**

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 (Abbildung 6). Ein Streudiagramm kommt an seine Grenze, sobald zu viele Observationen vorhanden sind (Abbildung 4). Man kann die Sättigung der Punkte minimieren, die Punkte verkleinern oder in unterschiedlichen Kategorien aufteilen, um eine bessere Übersicht zu erhalten (Abbildung 5). Empfehlenswert ist das Aggregieren der Daten, wobei bei der Aggregation eine Datenreduktion stattfindet und somit zusätzliche Informationen verloren gehen (Abbildung 6, 9). Verbindet man die Punkte miteinander, erhält man ein Liniendiagramm, auch Kurvendiagramm genannt (Abbildung 7, 8, 9). Ein Liniendiagram wird oft verwendet, um über einen bestimmten Zeitintervall eine Entwicklung oder Veränderung zu visualisieren oder festzustellen. Für die Visualisierung ist somit möglich, dass die x-Achse einen stetigen oder diskreten quantitativen sowie ordinale qualitative Daten sein kann. Ein Streudiagramm und Liniendiagramm ist eine einfache Visualisierung, jedoch muss man den Datensatz im Vorhinein verstehen und sich bewusst werden mit welchen Daten man arbeitet. In (Abbildung 4, 7) ist zu erkennen, dass zu viele Datenpunkte die Visualisierung unbrauchbar machen. Durch das Verständnis des Datensatzes, ist es erst möglich eine geeignete Visualisierung zu erstellen. In (Abbildung 6, 9) ist ersichtlich, dass man die Daten aggregieren muss, um eine geeignete Visualisierung zu finden. Somit konnte der Datensatz reduziert werden und die durchschnittliche Lebenserwartung pro Jahr geplottet werden.

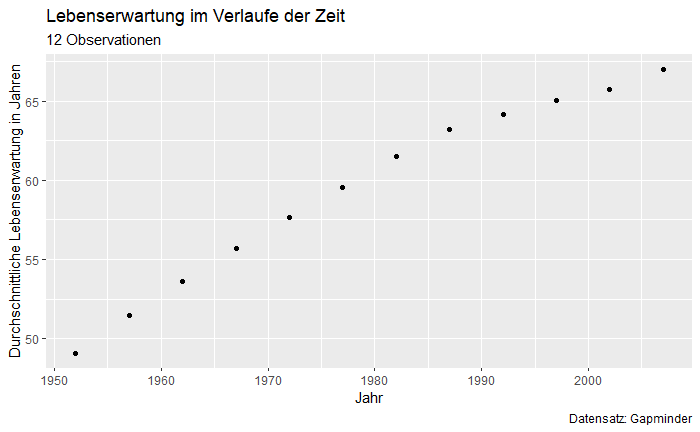
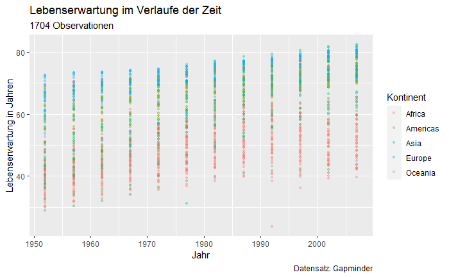
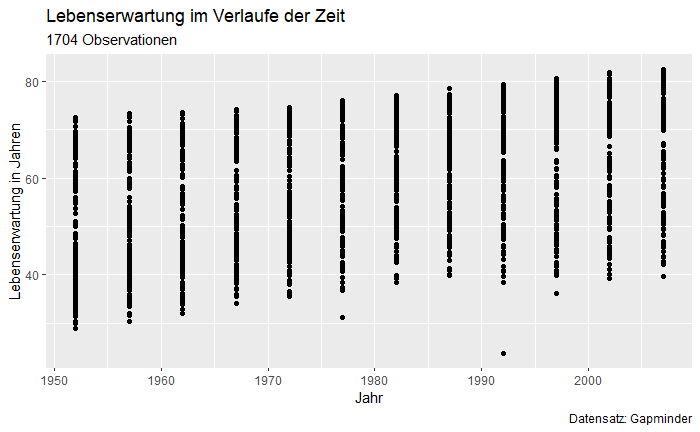


Abbildung 4 Punktediagramm Abbildung 5 nach Anpassung Abbildung 6 nach Aggregation

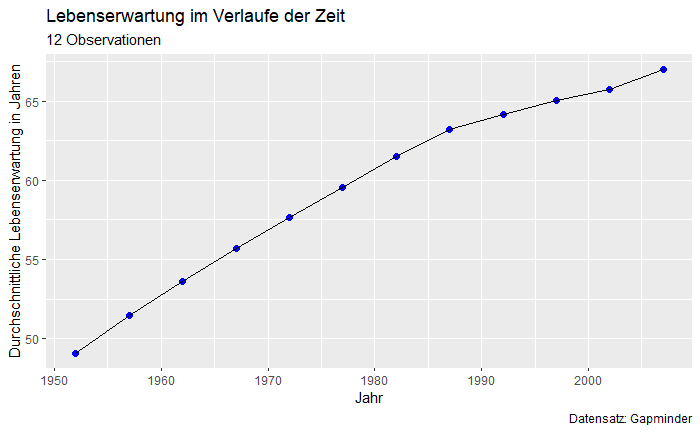
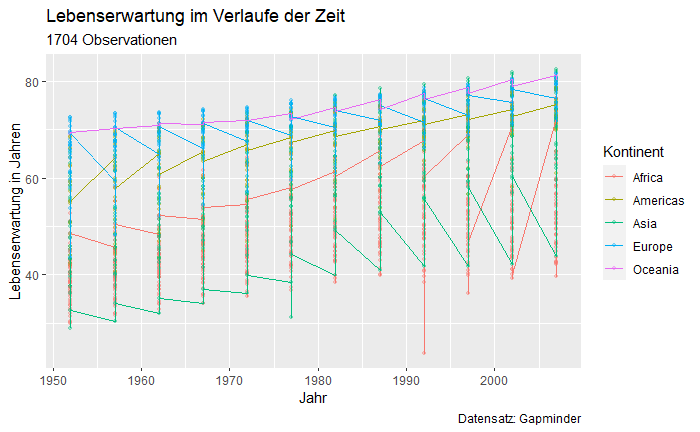
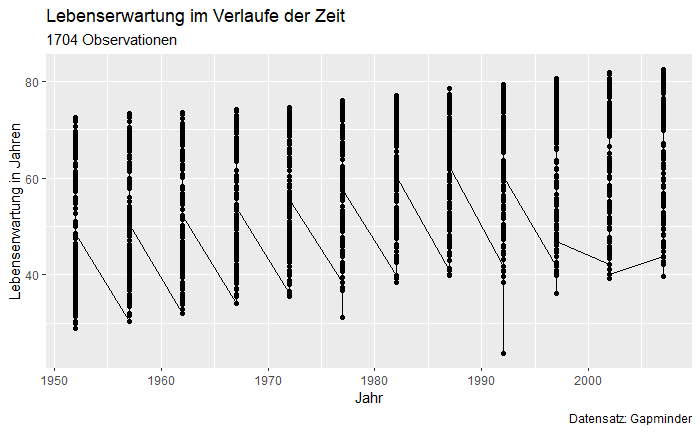


Abbildung 7 Liniendiagramm Abbildung 8 nach Anpassung Abbildung 9 nach Aggregation

## **Histogramme**

Histogramm ist eine Visualisierungen, 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. Somit ist die Visualisierung eines Histogramm von der Anzahl der Klassenbreite abhängig. Bei einer kleinen Anzahl der Klassenbreite, werden die Merkmale der Verteilung nicht angezeigt und gehen somit verloren (Abbildung 9). Auf der anderen Seite, wird bei einer grossen Anzahl Klassenbreite das Histogramm unübersichtlich und die Haupttrends der Daten können verloren gehen (Abbildung 10). Bei einem Histogramm muss man verschiedene Anzahl Klassenbreite überprüfen, um eine optimale Visualisierung zu finden, welches die Daten wiederspiegelt (Abbildung 12).

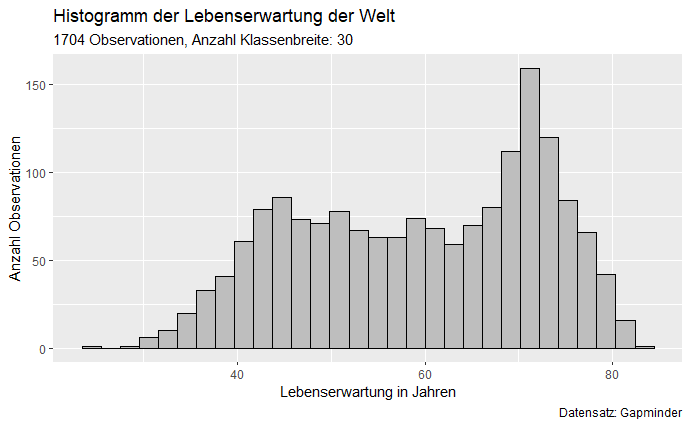
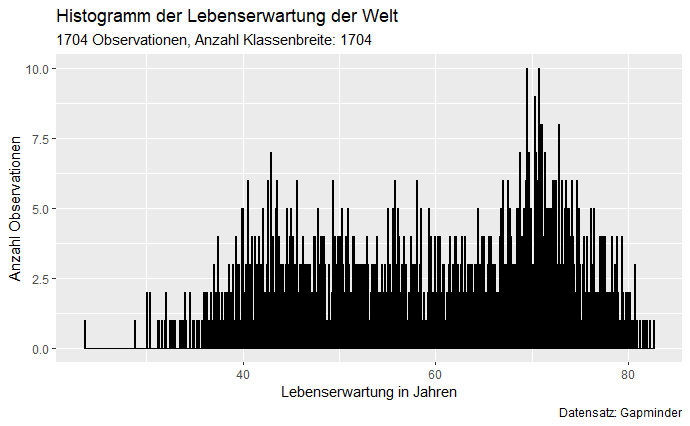
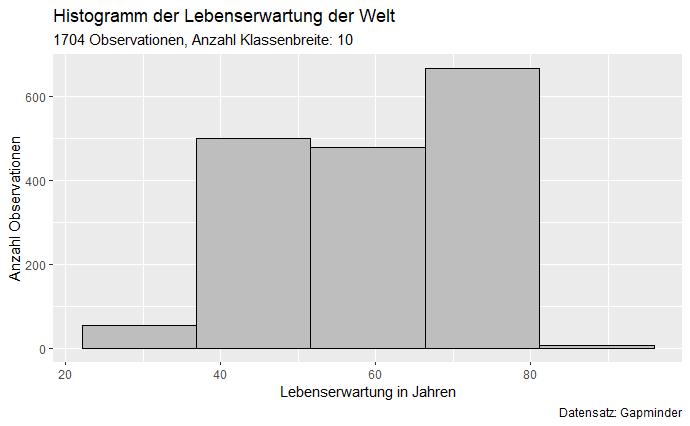


Abbildung 10 Anzahl bins = 5 Abbildung 11 Anzahl bins = 1704 Abbildung 12 Anzahl bins = 30

Ein Histogramm kann in Qualitativen Daten unterteilen werden. Dabei zu beachten ist, dass alle Kategorischen Daten unterschiedliche Farbe haben und voneinander klar trennbar sind. Auch muss man sich bewusst sein, mit welcher Art der Visualisierung man konfrontiert ist. In (Abbildung 13) sind die Histogramme aufeinander gestapelt. Das Interpretieren solcher Histogramme ist sehr schwierig. Einfacher sieht es bei der (Abbildung 14) aus. In dieser Visualisierung ist das Interpretieren einfacher, weil die Ausprägung entsprechend kategorischen Daten und Überlappung visualisiert sind. Bei der (Abbildung 15) ist die Überlappung nicht vorhanden, jedoch die Ausprägungen der einzelnen kategorische Werte.

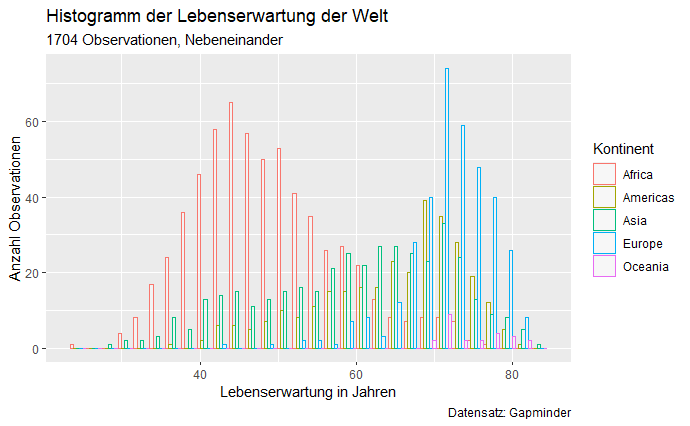
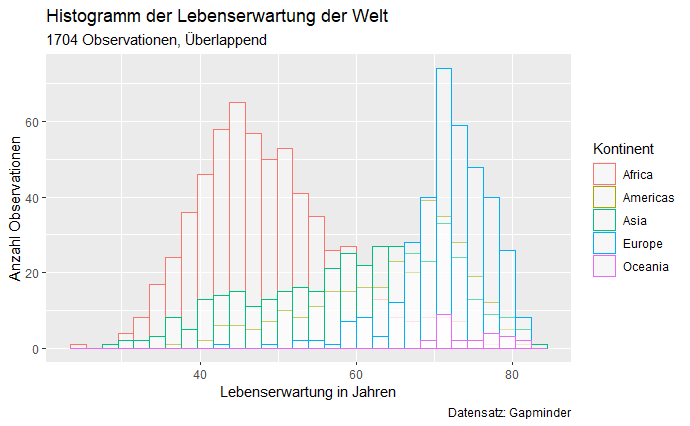
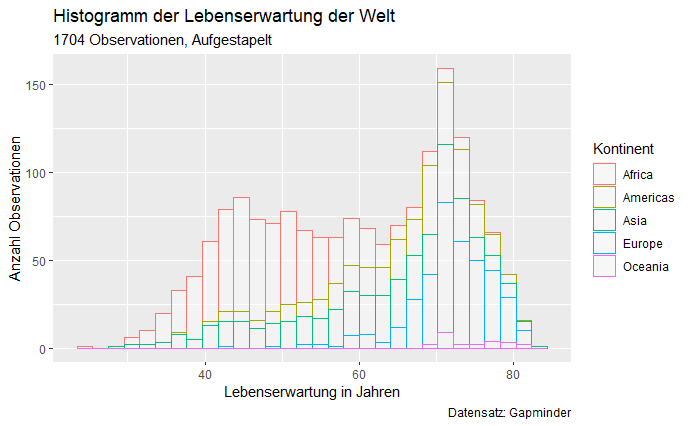


Abbildung 13 Überlappend Abbildung 14 Aufgestapelt Abbildung 15 Nebeneinander

# LO2: Visual Perception

# LO3: Design Principles vs. Data

# LO4: Grammar of Graphics Tools

# LO5: Evaluation

# Anhang Abbildungen

Abbildung: 1

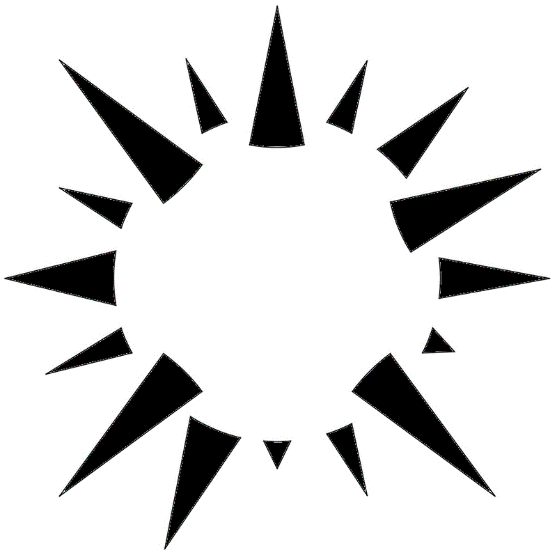


Abbildung: 2

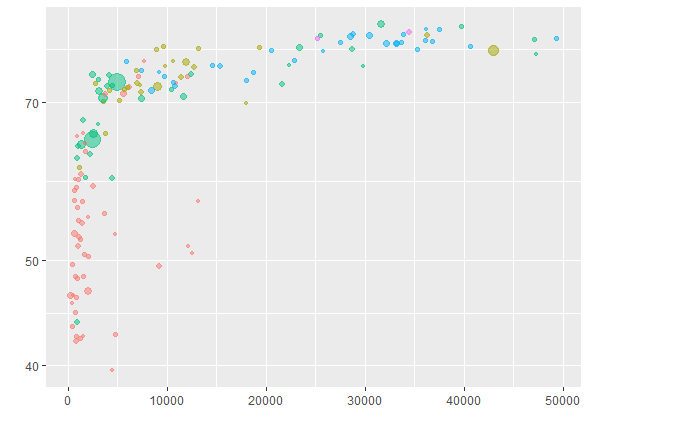


Abbildung: 3

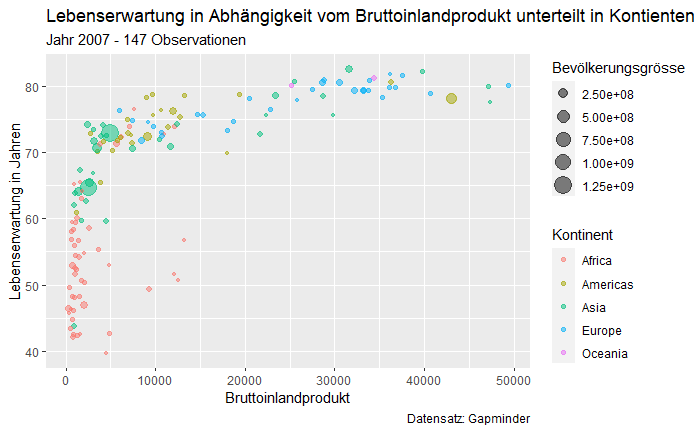


Abbildung: 4

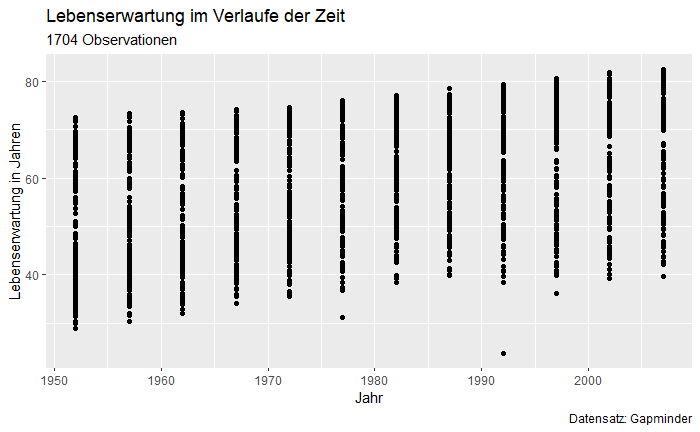


Abbildung: 5

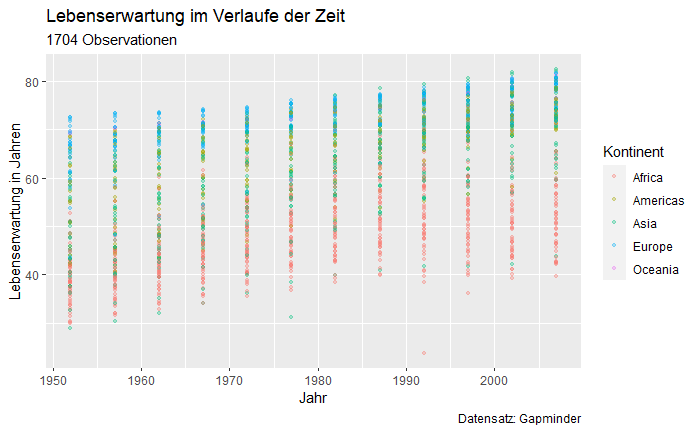


Abbildung: 6

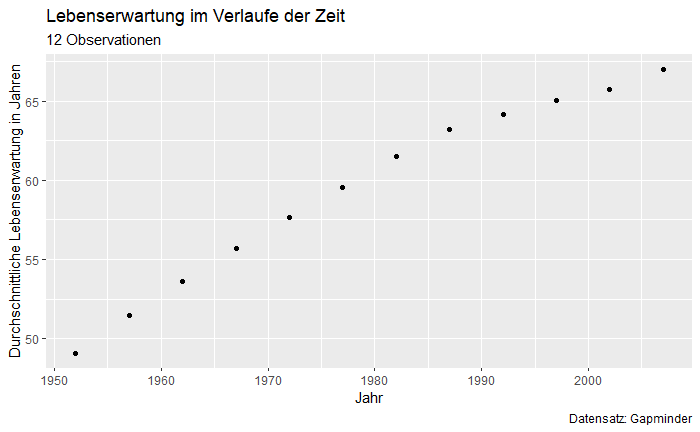


Abbildung: 7

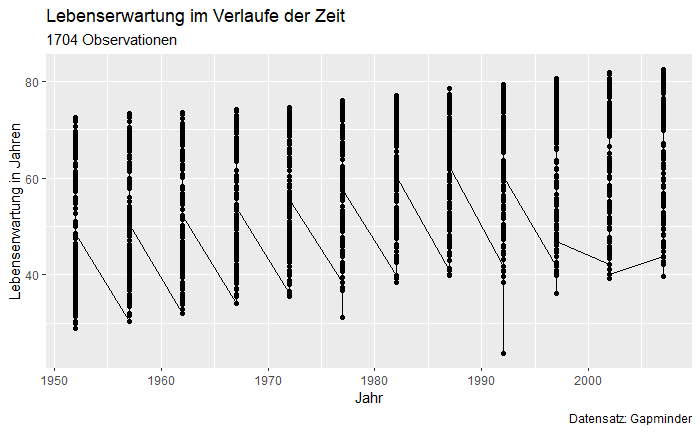


Abbildung: 8

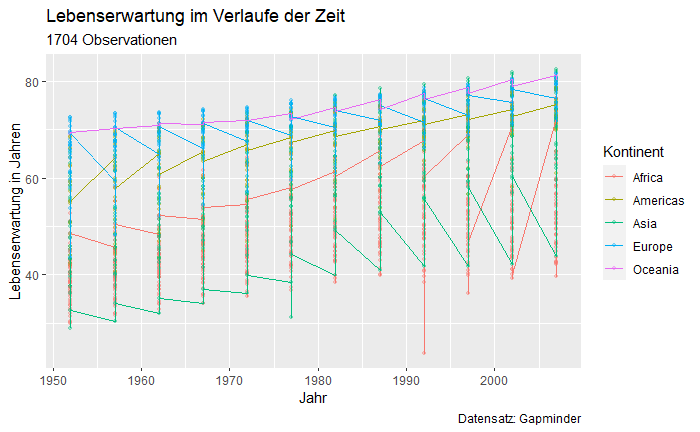


Abbildung: 9

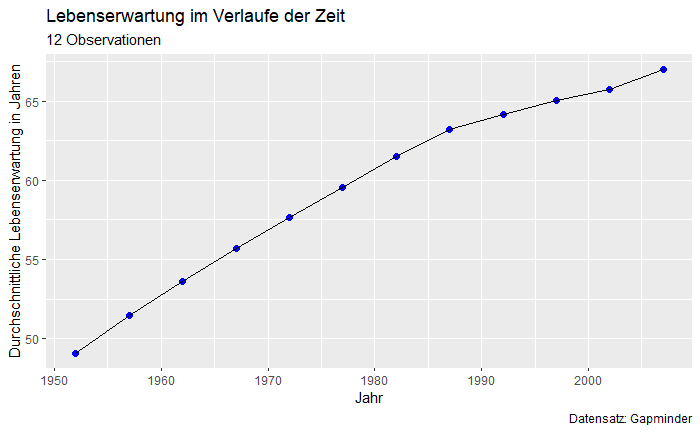


Abbildung: 10

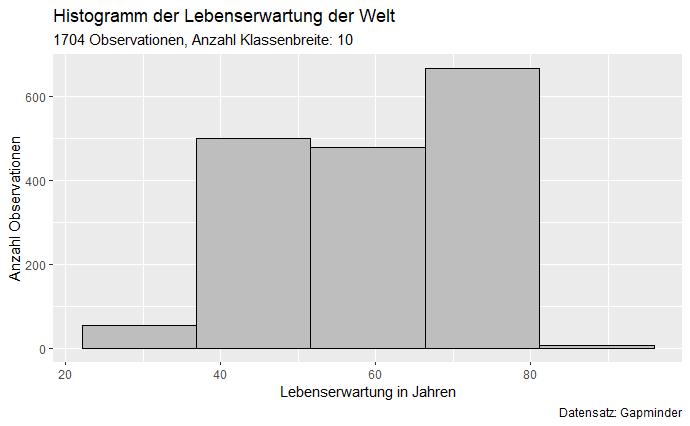


Abbildung: 11

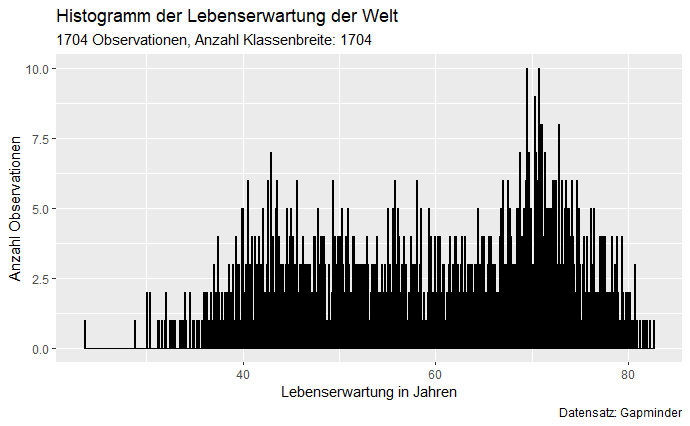


Abbildung: 12

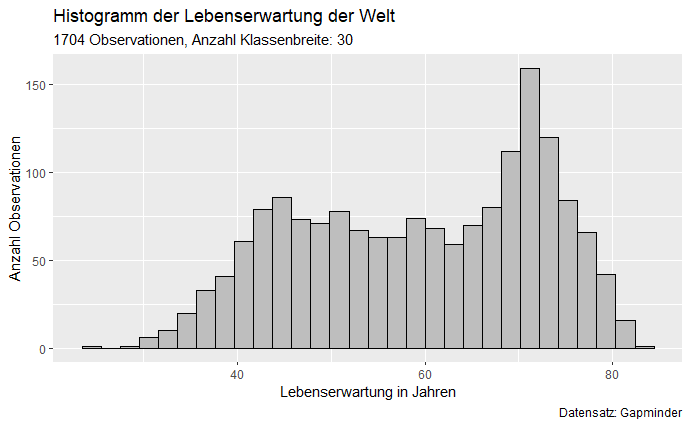


Abbildung: 13

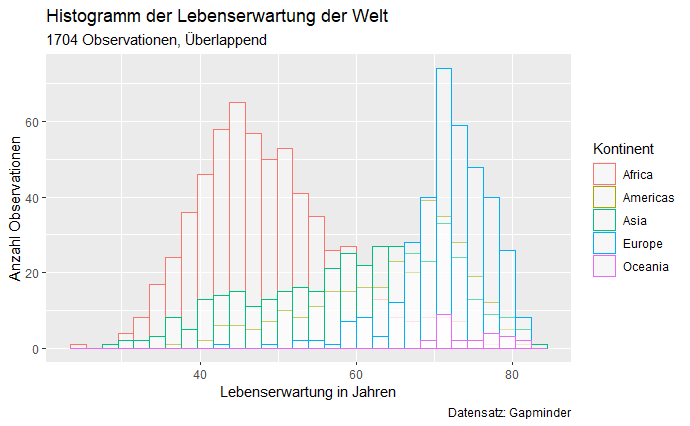


Abbildung: 14

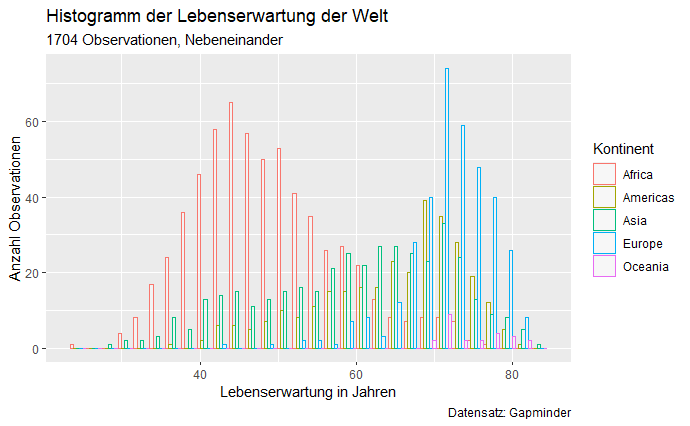


Abbildung: 15

