

Assignment im Fach
Data Visualization
bei Herrn Eichin

Abgabe am 20.05.2024

Von:

Aris Hofmann

Matrikelnummer:3979476

Fakultät Wirtschaft

Studiengang Wirtschaftsinformatik – Data Science

Kurs: WWI2023F

Data-Visualization-Dokumentation

Inhaltsverzeichnis

1. Einleitung	3
1.1. Wahl des Themas	3
2. Arbeit mit dem Datensatz	3
3. Erläuterung der Visualisierungen	4
3.1. Kategorisierte Produktionsmengen	4
3.2. Nahrungs- vs. Futtermittelproduktion	5
3.3. Interaktive Karte	5
3.4. Entwicklung der Nahrungs- und Futtermittelproduktion	6
3.5. In welchem Land gibt es die größten Produktionsveränderungen?	7
4. Grafiken	7
4.1 Kategorisierte Produktionsmengen	7
4.2 Nahrungsmittel- vs. Futtermittelproduktion	8
4.3 Interaktive Weltkarte	8
4.4 Zeitliche Veränderungen der Produktionsmengen	8
4.5 In welchem Land gibt es die größten Produktionsänderungen?	9
5. Literaturverzeichnis	9

1. Einleitung

Im Rahmen der Vorlesung ›Data Visualization‹ wurde uns die Abschlussaufgabe gestellt, durch Analyse eines Datensatzes fünf anschauliche Visualisierungen zu erstellen, wobei wir den Datensatz selbst wählen können.

Bei der Vorgehensweise sind die in der Vorlesung vorgestellten vier Schritte des Data-Visualization-Prozesses zu beachten, welcher aus zwei Hauptprozessen besteht. Der erste Hauptprozess gliedert sich in drei Schritte: Formulating your Brief, Working with Data, Establish editorial Thinking. Es geht dabei um die Planung, die Sammlung von Daten und die Erstellung der Story. Der zweite Hauptprozess (Develop Design Solution) setzt sich mit der konkreten Umsetzung der Ideen aus den ersten drei Teilschritten auseinander.

1.1. Wahl des Themas

Wir leben in einer Gesellschaft, in der sich die Mehrheit der Bevölkerung wenig Gedanken darüber macht, wo unser Essen herkommt. Wenn der Kühlschrank leer ist, gehen wir einfach zum Supermarkt und kaufen nach Lust und Laune ein. Uns stehen hunderte, wenn nicht sogar tausende fertige Produkte zur Verfügung, die sich ganz mühelos erwerben lassen.

Doch manchmal passieren unvorhergesehene Dinge. Beispielweise bricht in der Ukraine ein Krieg aus. Auf einmal werden alltägliche Produkte wie Sonnenblumenöl oder Getreide zu Mangelwaren und kosten den dreifachen Preis.

Um herauszufinden, woher die Lebensmittel eigentlich kommen, muss man hinter die Kulissen schauen. Die Lebensmittel tauchen nicht von allein in unseren Supermärkten auf. Ganz im Gegenteil: Sie werden größtenteils in Ländern in weiter entfernten Ländern der Welt hergestellt und uns über komplexe Logistikverfahren, von denen wir kaum etwas mitbekommen, zur Verfügung gestellt.

Die hier präsentierten Visualisierungen sollen dazu dienen, dem breiten Publikum eine bessere Vorstellung davon zu verschaffen, was welche Länder in welcher Menge an Nahrungsmitteln produzieren über einen bestimmten Zeitraum.

2. Arbeit mit dem Datensatz

Der gewählte Datensatz wurde auf der Plattform Kaggle gefunden und bezieht sich auf die Nahrungsmittel-Produktion aller Länder zwischen 1961 und 2013.¹ Die Daten stammen von der *Food and Agriculture Organization* (FAO) der Vereinten Nationen.² Hierbei handelt es sich um eine datentechnisch bereits gesäuberte Variante, die von einem israelischen Masterstudenten im Jahr 2018 hochgeladen wurde. Seinen Angaben nach war der Originaldatensatz

¹ <https://www.kaggle.com/dorbicycle/world-foodfeed-production>

² <https://www.fao.org/faostat/en/#home>

bereits so gut wie sauber, so dass nur noch geringfügige Änderungen vorgenommen zu werden brauchten, was hier vor Beginn der Arbeit an diesem Datensatz auch verifiziert wurde.

Die Daten werden als csv-Datei zur Verfügung gestellt und bestehen aus 63 Spalten, eingeteilt in „Area Abbreviation“, „Area Code“, „Area“, „Item Code“, „Item“, „Element“, „Unit“, „latitude“, „longitude“ und den jeweiligen Jahreszahlen von 1961 bis 2013. Bei „Element“ wird noch dazu zwischen menschlicher und tierischer Nahrung unterschieden.

Länder wie Grönland, Libyen, Papua Neu Guinea und die Demokratische Republik Kongo sind nicht mit enthalten. Der Datensatz zeigt keine Einheitsfehler und fehlende Werte treten nur auf bei Ländern, die zu diesem Zeitpunkt noch nicht gegründet waren.

Für die hier angefertigten Visualisierungen wurden die Spalten „latitude“ und „longitude“ entfernt, da sie für die Grafiken keine Rolle spielen.

Die Graphiken bauen aufeinander auf. Die ersten drei konzentrieren sich auf die Produktverteilung und veranschaulichen, aus welchen Ländern welche Produkte kommen. Diese Grafiken helfen dabei, die geografischen Unterschiede und Spezialisierungen zu erkennen. Die letzten beiden Grafiken wechseln zu einer zeitlichen Perspektive und zeigen anhand von Liniendiagrammen, wie sich die Produktionsmenge chronologisch entwickelt. Diese Darstellung ermöglicht es dem Betrachter, Tendenzen und Dynamiken in der Produktion zu erkennen. Somit wird ein grundlegendes Verständnis der aktuellen Produktionsverteilung geschaffen, bevor dann historische Entwicklungen und zeitliche Trends untersucht werden. Diese Herangehensweise stellt sicher, dass der Betrachter sowohl die gegenwärtige Situation als auch die historischen Veränderungen nachvollziehen kann, was zu einem tieferen Verständnis der Produktionsmuster führt.

Für die Visualisierungen wurde mit *Plotly* gearbeitet, das im Vergleich zu anderen Visualisierungstools, wie D3, deutlich benutzerfreundlicher ist und keine Vorkenntnisse in JavaScript verlangt. Plotly bietet eine Vielzahl an Möglichkeiten zur Darstellung von Daten, angefangen von einfachen Diagrammen bis hin zu komplexen interaktiven Grafiken. Um die Grafiken möglichst interaktiv zu gestalten, wurde *Dash* verwendet. Das ermöglicht Webanwendungen, in denen Benutzer auf einfache Weise interaktive Dashboards erstellen können. Ein großer Vorteil von Dash ist, dass es die sofortige Aktualisierung der Diagramme ermöglicht, was die Effizienz erheblich steigert, da zeitaufwändige manuelle Anpassungen entfallen.

3. Erläuterung der Visualisierungen

3.1. Kategorisierte Produktionsmengen

Das Ziel dieser Grafik ist es, die Produktionsmengen von verschiedenen Kategorien wie Getreide, Obst, Gemüse, Fleisch, Milchprodukte, Fisch, Öle und Fette, Getränke sowie andere Produkte zu analysieren und darzustellen. Durch die Auswahl verschiedener Länder, Jahre

und Visualisierungstypen können Trends und Unterschiede in der Produktion sichtbar gemacht werden. Diese interaktive Darstellung ermöglicht es den Nutzern, detaillierte Einblicke in die Verteilung der Produktion sowohl global als auch auf Länderebene zu gewinnen.

Für diese Analyse wurden zwei Visualisierungstypen gewählt: ein Grouped Bar Chart und ein Stacked Bar Chart. Diese Diagrammtypen eignen sich besonders gut, um die Produktionsmengen nach Kategorien darzustellen und zu vergleichen. Zusätzlich besteht die Möglichkeit, bestimmte Kategorien herauszufiltern, um die Darstellung weiter zu verfeinern.

Die Daten zeigen signifikante Unterschiede in den Produktionsmengen der verschiedenen Kategorien zwischen den Ländern. Durch Filter-Funktionen können auch kleinere Produktsegmente, wie zum Beispiel Fisch- und Ölprodukte, die leicht hinter den dominanten Hauptprodukten, wie Getreide- und Milchprodukte, verschwinden, Sichtbarkeit gewinnen.

Einige Länder spezialisieren sich aufgrund ihrer natürlichen klimatischen und landschaftlichen Bedingungen und landwirtschaftlichen Gegebenheiten und Traditionen auf bestimmte Produkte.

Durch die Auswahl verschiedener Jahre kann beobachtet werden, wie sich die Produktionsmengen in den einzelnen Kategorien im Laufe der Zeit verändert haben. Technologische Fortschritte und Änderungen in der Landwirtschaftspolitik können so nachvollzogen werden.

3.2. Nahrungs- vs. Futtermittelproduktion

Um die prozentuale Verteilung von menschlicher und tierischer Nahrungsmittelproduktion bestmöglich darzustellen, wurde die Form des Kreisdiagramms gewählt. Benutzer können über Dropdown-Menüs das gewünschte Land und Jahr auswählen.

Die Daten zeigen, dass in vielen Ländern die Produktion von Nahrungsmitteln (Food) dominiert. Einige Länder hingegen weisen eine signifikante Produktion von Futtermitteln (Feed) auf, was auf unterschiedliche landwirtschaftliche Schwerpunkte und wirtschaftliche Strukturen hinweist.

Durch die Auswahl verschiedener Jahre lässt sich gut verfolgen, wie sich die Produktionsmengen im Laufe der Zeit verändert haben.

Die Visualisierung ermöglicht es, spezifische Länder detailliert zu betrachten. Länder mit stark variierender Produktion können so identifiziert werden als Basis für eine über das vorliegende Projekt hinausgehende Erforschung der Gründe für diese Schwankungen.

3.3. Interaktive Karte

Die interaktive Karte verwendet Dropdown-Menüs, um das gewünschte Item, das Element und das Start- bzw. Endjahr auszuwählen. Durch diese Auswahlmöglichkeiten können gezielt Daten gefiltert und spezifische Analysen durchgeführt werden.

Neben der reinen Produktionsmenge wird auch die Verfügbarkeit und die Historie der Daten verdeutlicht. Länder, für die keine Daten vorhanden sind, werden explizit gekennzeichnet. Ebenso wird angezeigt, wenn ein Land zu einem bestimmten Zeitpunkt noch nicht gegründet war. Dies bietet eine zusätzliche Dimension der Kontextualisierung und historischen Einordnung.

Die Karte zeigt deutlich die regionalen Unterschiede in der Produktionsmenge. Länder wie China und die USA weisen, wie oben bereits erwähnt, hohe Produktionsmengen auf, während andere Regionen wie Afrika geringere Mengen produzieren.

Die Auswahl größerer Zeiträume ermöglicht es, langfristige Trends zu erkennen: wie sich die Produktionsmengen global verschieben, welche Länder zu Hauptproduzenten aufsteigen und welche Länder im internationalen Ranking an Bedeutung verlieren.

Die Möglichkeit, zwischen verschiedenen Items (z.B. Weizen, Reis, Mais) und Elementen (Food oder Feed) zu wählen, erlaubt eine detaillierte Analyse der Produktionsketten. So kann beispielsweise die Bedeutung von Getreide für die Nahrungsmittelproduktion im Vergleich zur Futtermittelproduktion untersucht werden.

Länder mit höheren Produktionsmengen werden in dunkleren Rot-Tönen dargestellt, während Länder mit geringeren Mengen in helleren Farben. Länder ohne Daten sind in Weiß angezeigt. Diese Farbskala ermöglicht es, die Produktionsmengen auf einen Blick zu erfassen und Unterschiede leicht zu erkennen.

3.4. Entwicklung der Nahrungs- und Futtermittelproduktion

Die zeitliche Entwicklung der Produktionsmengen wird anhand eines Liniendiagramms dargestellt. Über Dropdown-Menüs können die Nutzer zwischen den Elementen „Food“ (Essen) und „Feed“ (Futter) wählen sowie einzelne oder mehrere Länder auswählen, um deren Produktionsverläufe zu vergleichen. Die Auswahl „all countries“ zeigt die Produktionsdaten aller Länder gleichzeitig an.

Durch die Zusammenstellung der Daten nach Jahr und Land werden die jährlichen Gesamtproduktionsmengen berechnet und dargestellt. Die Daten umfassen die Jahre ab 1961 und bieten somit einen umfassenden historischen Überblick.

Länder wie die USA und China zeigen über die Jahre hinweg beträchtliche Produktionssteigerungen, was auf technologische Fortschritte und verbesserte landwirtschaftliche Produktionsprozesse zurückzuführen sein dürfte.

Bei einigen Ländern gibt es größere Schwankungen, die in der Regel mit wirtschaftlichen oder politischen Krisen oder Naturkatastrophen zusammenhängen.

Die Liniendiagramme sind so gestaltet, dass sie sowohl die Gesamtproduktion als auch spezifische Trends innerhalb der Länder deutlich abbilden, wobei die farbliche Gestaltung dabei hilft, die Unterschiede zwischen den Ländern und Elementen hervorzuheben.

3.5. In welchem Land gibt es die größten Produktionsveränderungen?

Anhand der prozentualen Veränderung des Gesamtwerts von einem Jahr zum nächsten, wurden bei dieser Analyse die jährlichen Änderungsraten der Produktion berechnet. Dies ermöglicht es, nicht nur absolute Produktionsmengen, sondern auch relative Veränderungen zu betrachten. Das Land mit den weltweit höchsten jährlichen Produktionssteigerungen oder -rückgängen wird angezeigt.

Es wurde bewusst darauf verzichtet, konkrete Ländernamen direkt in der Grafik anzugeben, um die Lesbarkeit zu erleichtern. Die einzelnen Ländernamen erscheinen, wenn man mit der Maus über die graphischen Punkte im Diagramm fährt (Hovern). Diese Methode unterstützt eine saubere und übersichtliche Darstellung, ohne die Betrachter mit zu vielen Details zu überfrachten.

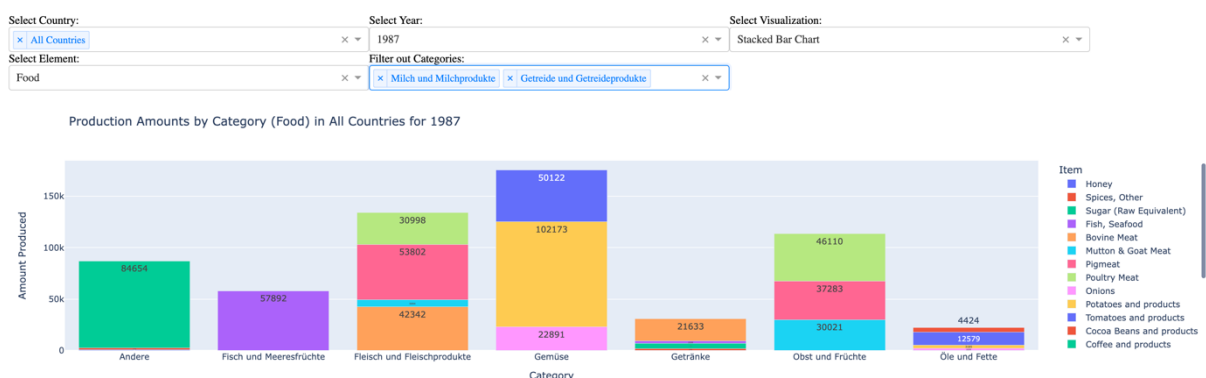
Die Wahl des Liniendiagramms für diese Analyse wurde getroffen, um die Klarheit und Verständlichkeit der Daten zu maximieren. Im Vergleich zu anderen Diagrammtypen wie Balken- oder Streudiagrammen bietet das Liniendiagramm Chart eine prägnantere Darstellung von Spitzenwerten, ohne die Übersichtlichkeit zu beeinträchtigen. Dies ist besonders wichtig, wenn es darum geht, extreme Werte hervorzuheben und gleichzeitig einen klaren Überblick über die Daten zu behalten.

Die Visualisierung zeigt auch, dass die Produktionsänderungen nicht gleichmäßig über alle Regionen verteilt sind. Tendenziell zeigen bestimmte Regionen wie Asien und Südamerika größere Schwankungen als andere, was auf regionale oder wirtschaftliche Entwicklungen zurückgeführt werden kann.

4. Grafiken

4.1 Kategorisierte Produktionsmengen

Production Amounts by Category for a Single Year



4.2 Nahrungsmittel- vs. Futtermittelproduktion

FAO Data Visualization

Select Country: x Select Year: x

Food vs Feed Production for Germany in 2011

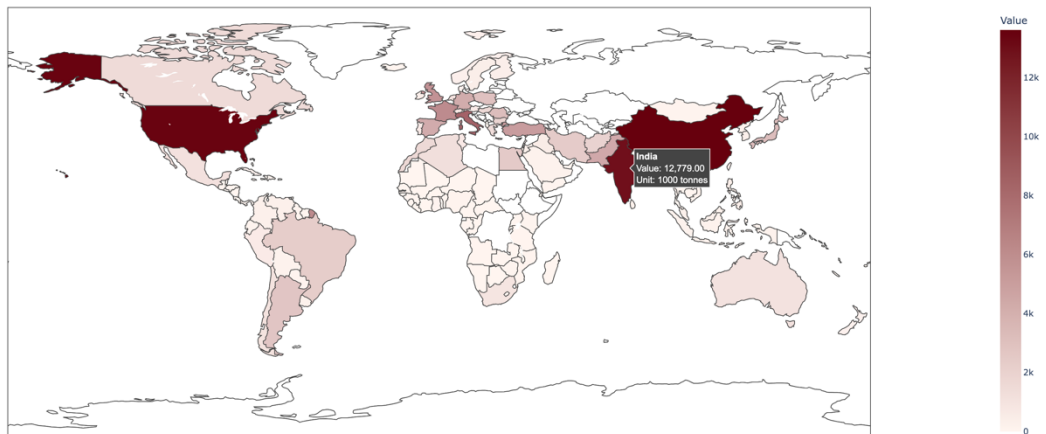


4.3 Interaktive Weltkarte

World Production Map

Select Item: x Select Year: x Select Element: x

Food Production for Wheat and products (1961)

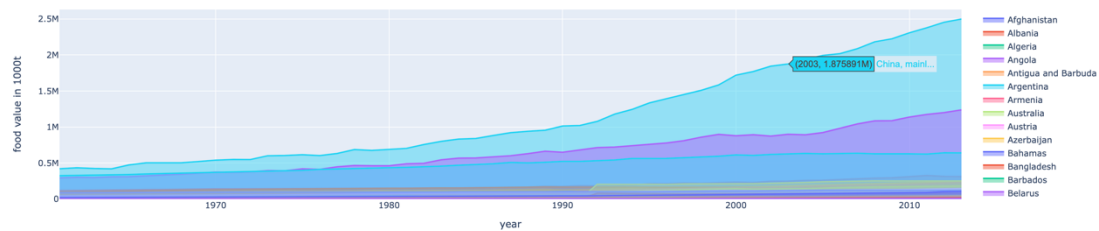


4.4 Zeitliche Veränderungen der Produktionsmengen

Development Of Production Over Time

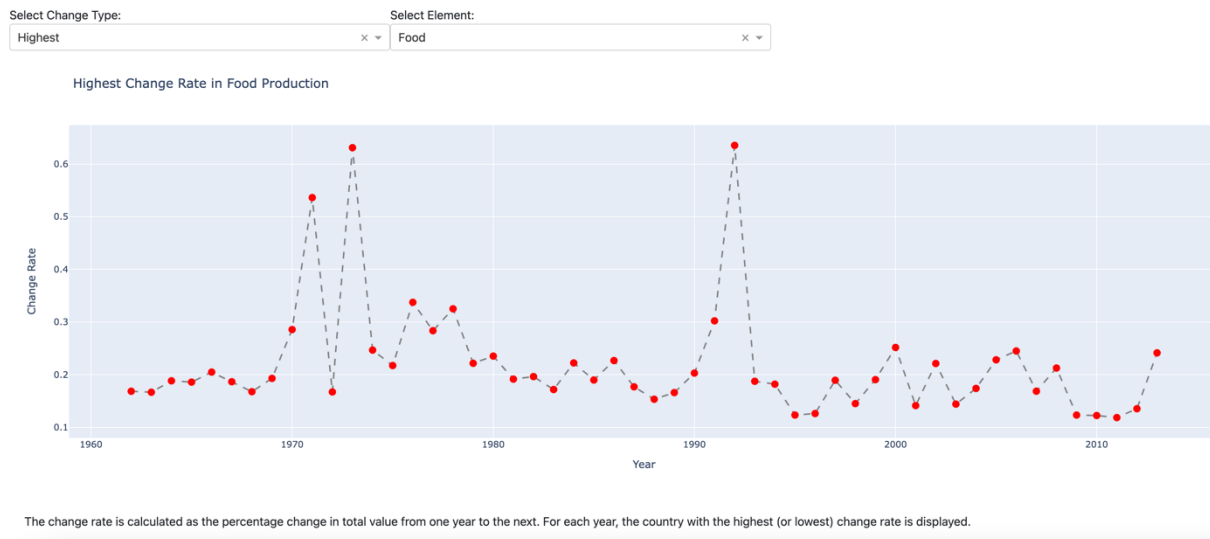
Select Element: x Select Countries: x

Food time series per country



4.5 In welchem Land gibt es die größten Produktionsänderungen?

Line Chart of Change Rates



5. Literaturverzeichnis

Dash Documentation & User Guide | Plotly (Dash Documentation & User Guide | Plotly, 2023), <https://dash.plotly.com/> (2024-05-06) [Zugriff 2024-05-11]

Plotly Documentation | Plotly (Plotly Documentation | Plotly, 2023), <https://plotly.com/python/> (2024-05-06) [Zugriff 2024-05-06]

Pandas Documentation | Pandas (Pandas Documentation | Pandas, 2023), <https://pandas.pydata.org/pandas-docs/stable/> (2024-05-06) [Zugriff 2024-05-06]

Pycountry Documentation | PyPI (Pycountry Documentation | PyPI, 2023), <https://pypi.org/project/pycountry/> (2024-05-06) [Zugriff 2024-05-09]

FAO Statistics Database | FAO (FAO Statistics Database | FAO, 2023), <http://www.fao.org/fao-stat/en/> (2024-05-06) [Zugriff 2024-05-04]

Stack Overflow (Stack Overflow, 2023), <https://stackoverflow.com/> (2024-05-06) [Zugriff 2024-05-12]

GeeksforGeeks (GeeksforGeeks, 2023), <https://www.geeksforgeeks.org/> (2024-05-06) [Zugriff 2024-05-012]

Real Python (Real Python, 2023), <https://realpython.com/> (2024-05-06) [Zugriff 2024-05-012]

Towards Data Science | Medium (Towards Data Science | Medium, 2023), <https://towardsdatascience.com/> (2024-05-06) [Zugriff 2024-05-014]

Tutorialspoint (Tutorialspoint, 2023), <https://www.tutorialspoint.com/> (2024-05-06) [Zugriff 2024-05-06]

Kaggle (Kaggle, 2023), <https://www.kaggle.com/> (2024-05-06) [Zugriff 2024-05-01]