



# NameStats

## Datenanalyse amerikanischer Vornamen

Exposee

Ein Programm für die Analyse und Auswertung von Vornamen in den USA

Alexander Stach

<https://github.com/AudioSonic/NameStats>

## Inhaltsverzeichnis

Einleitung.....	2
Funktionen.....	3
Statistik.....	3
Vornamen.....	3
Geschlechter.....	3
Bundesstaaten.....	3
Ausgabe .....	3
Planung.....	4
Das Interface .....	4
Programmablaufplan.....	5
Umsetzung.....	6
Hauptkomponenten .....	6
Benutzeroberfläche (GUI).....	6
Datenverarbeitung .....	6
Grafische Darstellung .....	6
Funktionen.....	7
eingabe().....	7
staaten_woerterbuch.....	7
Benutzeroberfläche .....	7
Problembehandlung.....	8

## Einleitung

NameStats ist ein interaktives Analysewerkzeug, die es Benutzern ermöglicht, Statistiken über Vornamen in den USA abzurufen und zu analysieren. Mit NameStats können Benutzer einen Vornamen eingeben und erhalten detaillierte Informationen darüber, wie häufig dieser Name im Laufe der Zeit in den USA vertreten war.

Die Anwendung bietet auch die Möglichkeit, die Suche durch die Angabe von zusätzlichen Parametern einzuschränken. Benutzer können das Geschlecht des Vornamens spezifizieren, einen bestimmten Bundesstaat auswählen oder eine bestimmte Zeitspanne festlegen, um genaue und maßgeschneiderte Statistiken zu erhalten.

Ein Merkmal von NameStats ist die grafische Darstellung der Veränderung des Namens über die Zeit. Durch die Verwendung von Graphen können Benutzer leicht nachvollziehen, wie die Beliebtheit eines bestimmten Namens im Laufe der Jahre gestiegen oder gesunken ist.

Die Anwendung wurde unter Verwendung von Visual Studio Code und der Programmiersprache Python entwickelt. Um auf die erforderlichen Daten zuzugreifen, wurde eine CSV-Datei verwendet, die umfassende Informationen über Vornamen in den USA enthält. NameStats bietet eine benutzerfreundliche und informative Möglichkeit, Einblicke in die Entwicklung von Vornamen im Laufe der Zeit zu gewinnen.

## Funktionen

### Statistik

NameStats bietet eine detaillierte Statistik über Vornamen in den USA, die den Zeitraum von 1910 bis 2014 abdeckt. Benutzer haben die Möglichkeit, die Zeitspanne der Statistik mit einem Schieberegler individuell anzupassen, um spezifische Jahrzehnte oder Zeiträume zu untersuchen.

### Vornamen

Die Datenbank von NameStats enthält eine umfangreiche Liste von Vornamen, die im Laufe der Jahre in den USA verwendet wurden. Diese Liste ist äußerst umfassend und beinhaltet sowohl gebräuchliche als auch seltene Namen. Benutzer können Informationen über nahezu jeden erdenklichen Vornamen erhalten und seine Popularität im Laufe der Zeit verfolgen.

### Geschlechter

Jedem Namen in der Datenbank ist ein Geschlecht (Männlich oder Weiblich) zugeordnet. Da viele Vornamen in beiden Geschlechtskategorien auftreten können, bietet NameStats Filteroptionen, um Benutzern die Auswahl des gewünschten Geschlechts zu ermöglichen. Dies ermöglicht es Benutzern, Statistiken für männliche, weibliche oder geschlechtsneutrale Vornamen zu generieren.

### Bundesstaaten

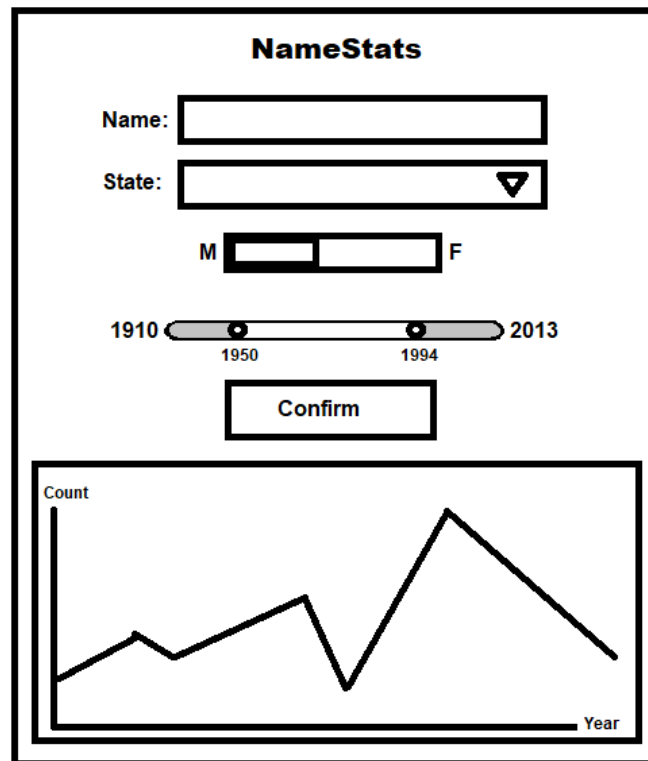
Zusätzlich zur Filterung nach Geschlecht können Benutzer die Statistiken auch nach Bundesstaaten aufschlüsseln. NameStats enthält eine vollständige Liste aller Bundesstaaten der USA, einschließlich ihrer Kürzel (z.B., AK für Alaska oder NY für New York). Diese Funktion ermöglicht es Benutzern, regionale Unterschiede in der Verwendung von Vornamen zu erkunden und die Beliebtheit eines Namens in einem bestimmten Bundesstaat zu überprüfen.

### Ausgabe

Je nach Eingabe des Benutzers werden verschiedene Werte in Form einer Textbox ausgegeben. Sollte der Benutzer fehlerhafte Daten eingetragen haben (wie beispielsweise ein größeres Mindestjahr als das Maximaljahr) so wird es ihm entsprechend mitgeteilt. Falls die Eingaben korrekt sind, werden ihm folgende Informationen ausgegeben: Der maximale und minimale Wert an Personen, die in dem angegebenen Zeitraum mit dem vom Benutzer bestimmten Namen versehen wurden sowie jedes Jahr mit der jeweiligen Anzahl an Personen pro Jahr.

## Planung

### Das Interface



The diagram shows a web interface titled "NameStats". It contains the following elements:

- Name:** A text input field.
- State:** A dropdown menu with a downward arrow.
- Gender:** A slider between "M" (Male) and "F" (Female). The slider is currently positioned towards "M".
- Time Range:** A slider between the years "1910" and "2013". There are two markers on the slider, one at "1950" and another at "1994".
- Confirm:** A button labeled "Confirm".
- Graph:** A line graph with "Count" on the y-axis and "Year" on the x-axis. The graph shows a fluctuating line with several peaks and troughs.

**Name:** Ein simples Textfeld, welches den jeweiligen Vornamen aufnimmt und auswertet. Ist ein Name vorhanden, so ergibt es einen Rückgabewert, falls nicht, dann soll eine entsprechende Fehlermeldung erscheinen.

**State:** Eine Auflistung aller Staaten in den USA. Hierbei handelt es sich um ein Dropdown Menü bei dem der Benutzer seinen gewünschten Staat aussuchen kann. Das verhindert doppelte Einträge (wie beispielsweise den Eintrag mehrerer Staaten) und erleichtert dem Benutzer die Auswahl, weil dieser die Staaten nicht aus dem Kopf können muss.

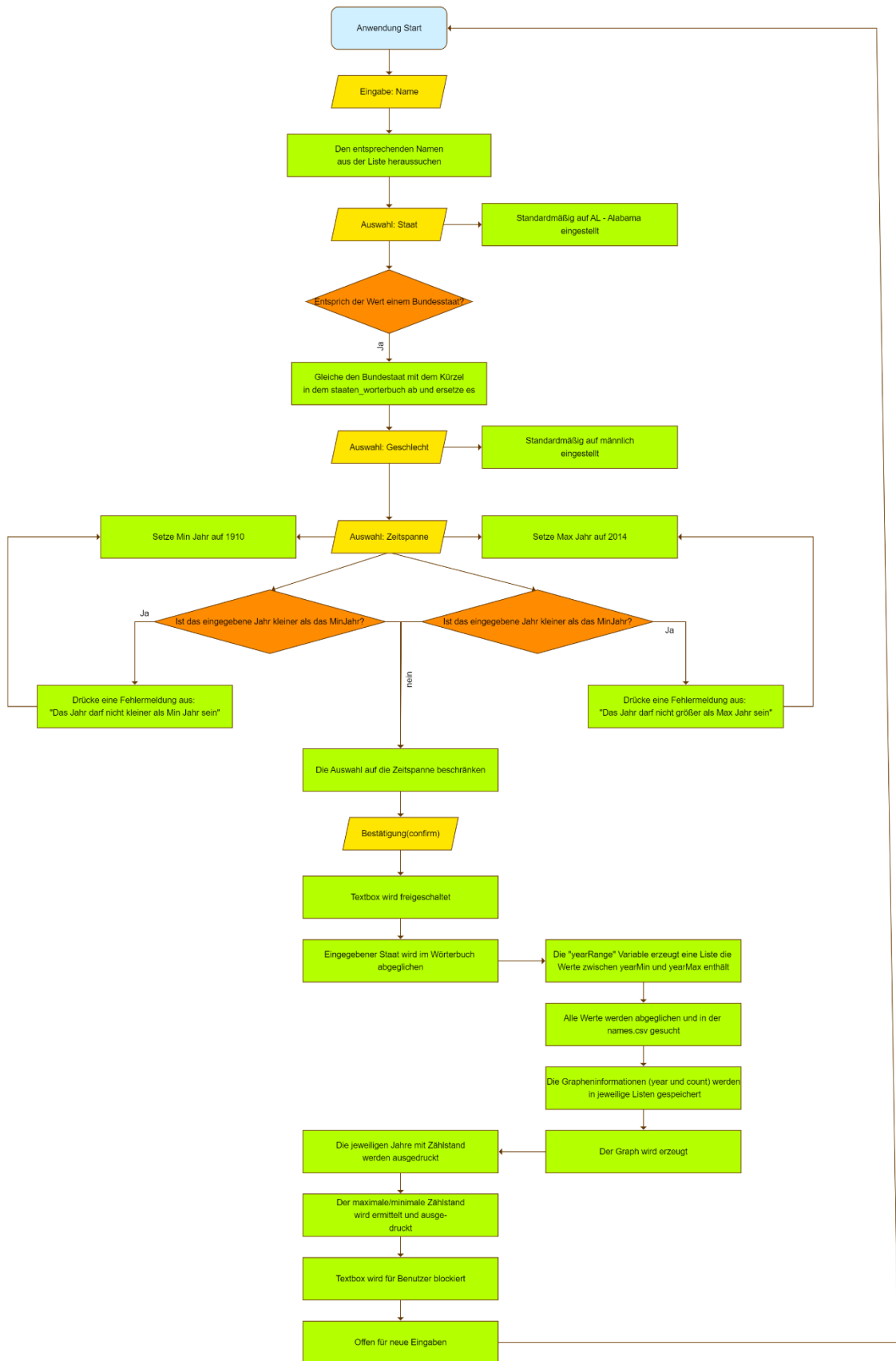
**Geschlecht:** Dies soll mithilfe eines Sliders geregelt werden. Der Wert 0 ergibt dabei männlich und der Wert 1 ergibt weiblich. Da die Statistik keine diversen Geschlechter enthält, ist es auch nicht notwendig weitere Optionen einzubauen. Standardmäßig ist der Slider auf männlich eingestellt.

**Zeitspanne:** Die maximal mögliche Spanne geht zwischen 1910 und 2013. Der Benutzer ist jedoch in der Lage die Spanne selber zu definieren und die Slider nach seinen Wünschen anzupassen.

**Confirm:** Ein simpler Button, welcher die Eingaben bestätigt und den Graphen erzeugt. Sollte ein fehlerhafter Name (also beispielsweise einer mit Sonderzeichen) eingetragen sein, dann wird eine Fehlermeldung erzeugt.

**Graph:** Der Graph wird mit Hilfe der matplotlib Bibliothek erzeugt. Die X-Achse stellt das Jahr und die Y-Achse die Häufigkeit der Namen in dem jeweiligen Jahr dar. Der Graph kann nur erzeugt werden, wenn alle Eingaben passen. Sollten die Eingaben passen, der Name in der Liste jedoch nicht vorkommen, dann wird der Graph einfach als flache Linie dargestellt.

# Programmablaufplan



## Umsetzung

Für die Umsetzung wurde das Programm Visual Studio 2022 verwendet. Es hat mir alle nötigen Funktionen geboten, um problemlos mit Python arbeiten zu können. Als Grundlage für das gesamte Programm habe ich eine .csv Datei verwendet, welche alle benötigten Informationen enthielt.

## Hauptkomponenten

Das Skript enthält die folgenden Hauptkomponenten:

### Benutzeroberfläche (GUI)

Das Programm verfügt über eine benutzerfreundliche grafische Benutzeroberfläche (GUI), die mithilfe der tkinter-Bibliothek erstellt wurde. Die GUI enthält verschiedene Eingabeelemente, darunter ein Textfeld für die Eingabe eines Namens, ein Dropdown-Menü zur Auswahl eines Bundesstaats, einen Slider zur Geschlechtsauswahl (mit den Werten 0 und 1) und zwei Spinboxen für die Auswahl eines Jahres.

Zusätzlich zur GUI nutzt das Programm die matplotlib-Bibliothek, um auf Grundlage vorher festgelegter Achsenwerte einen Graphen zu generieren. Dieser Graph kann verwendet werden, um Daten zu visualisieren und Trends darzustellen.

Ein zentraler "Submit"-Button in der Benutzeroberfläche löst die Analyse aus und ermöglicht es dem Benutzer, die gewünschten Informationen zu erhalten. Insgesamt bietet das Programm eine interaktive Möglichkeit, Daten zu analysieren und Ergebnisse grafisch darzustellen.

### Datenverarbeitung

Das Skript verarbeitet die Eingaben des Benutzers und führt eine Suche in der `names.csv`-Datei durch, um Statistiken zu Vornamen zu erstellen. Um sicherzustellen, dass die Benutzereingaben korrekt verarbeitet werden, wird eine Liste von Bundesstaaten verwendet, die mit einem Wörterbuch von Kürzeln verknüpft ist. Dies ist erforderlich, da die Staaten in der `names.csv`-Datei nur mit Kürzeln angegeben sind, während das Programm die vollen Namen der Staaten verwenden soll. Diese Verknüpfung stellt sicher, dass keine Verwirrung unter den Benutzern entsteht und sie nicht nach den entsprechenden Kürzeln suchen müssen. Dadurch wird die Benutzerfreundlichkeit und Klarheit der Benutzeroberfläche erhöht, da die Benutzer die Bundesstaaten leicht anhand ihrer vollen Namen auswählen können, ohne sich mit Kürzeln befassen zu müssen.

### Grafische Darstellung

Die Generierung von Statistiken und deren grafische Darstellung sind zentrale Funktionen dieses Skripts. Dies wird durch die Verwendung der matplotlib-Bibliothek ermöglicht, die es dem Programm möglich macht, Diagramme und Grafiken zu erstellen.

Die matplotlib-Bibliothek bietet eine umfangreiche Palette von Funktionen zur Visualisierung von Daten. In diesem Fall wird sie verwendet, um Diagramme basierend auf den in der names.csv-Datei gefundenen Informationen zu generieren. Diese Diagramme können verschiedene Statistiken darstellen, wie beispielsweise die Häufigkeit eines bestimmten Namens über die Jahre hinweg.

Durch die grafische Darstellung der Statistiken werden die Daten für den Benutzer leicht verständlich und ansprechend präsentiert. Dies ermöglicht es dem Benutzer, Trends und Muster auf einen Blick zu erkennen und die Daten effektiv zu interpretieren. Die Verwendung von matplotlib trägt erheblich zur Benutzerfreundlichkeit und Aussagekraft des Programms bei, indem sie die Informationen auf visuell ansprechende Weise präsentiert.

## Funktionen

### eingabe()

Die Hauptfunktion des Skripts wird aufgerufen, wenn der Benutzer auf den "Submit"-Button klickt. Diese Funktion verarbeitet die Benutzereingaben und führt die Analyse durch.

### Funktionsablauf

#### 1. Erfassung von Benutzereingaben:

- Der eingegebene Name wird aus dem Textfeld abgerufen.
- Der ausgewählte Bundesstaat wird aus dem Dropdown-Menü abgerufen.
- Das ausgewählte Geschlecht wird anhand des "gender" Sliders ermittelt.
- Die ausgewählte Zeitspanne wird durch zwei Spinboxen (min Jahr und max Jahr) erfasst.

#### 2. Umwandlung des Bundesstaates:

- Die Benutzereingabe für den Bundesstaat wird in das entsprechende Kürzel umgewandelt, indem das "staaten\_woerterbuch" verwendet wird.

#### 3. Datenverarbeitung:

- Das Skript durchläuft die "names.csv"-Datei und sucht nach Übereinstimmungen mit den Benutzereingaben.
- Wenn Übereinstimmungen gefunden werden (Name, Bundesstaat, Geschlecht und Jahr), werden die Statistiken angezeigt.

#### 4. Grafische Darstellung:

- Ein Diagramm mit den gefundenen Statistiken wird mithilfe von matplotlib erstellt und angezeigt.

### staaten\_woerterbuch

Das Staatenwörterbuch ist eine Datenstruktur, die die vollen Namen der US-Bundesstaaten mit ihren jeweiligen Kürzeln verknüpft. Dies dient dazu, die Benutzereingaben für Bundesstaaten zu normalisieren und sicherzustellen, dass das Programm die korrekten Abkürzungen verwendet, wenn es nach Bundesstaaten in der "names.csv"-Datei sucht

### Benutzeroberfläche

Die Benutzeroberfläche wurde mithilfe der tkinter-Bibliothek erstellt und enthält die folgenden Elemente:

- Ein Textfeld zur Eingabe des Namens.
- Ein Dropdown-Menü zur Auswahl des Bundesstaats.
- Ein Slider zur Geschlechtsauswahl (0 oder 1 - weiblich oder männlich).
- Zwei Spinboxen zur Auswahl der Jahrespanne
- Ein "Submit"-Button zur Ausführung der Analyse.



## Problembehandlung

Zunächst wurde das Programm Visual Studio Code fürs Programmieren in Erwägung gezogen. Es gab jedoch Probleme damit die .csv Datei zu lesen, weshalb ich mich doch eher für Visual Studio 2022 entschieden habe. Letzten Endes machte die Entwicklungsumgebung für mich keinen großen Unterschied, weshalb der Wechsel zu VS 2022 keine großen Nachteile mit sich zog.

Ursprünglich war es geplant ein Widget mit zwei Slidern für die Jahre zu verwenden. Tkinter scheint jedoch keine Möglichkeit zu haben mehr als einen Slider in einem Widget zu haben, weshalb ich mir etwas anderes überlegen musste. Als Alternative habe ich zwei Spinboxen genommen, welche mit einer Liste verbunden waren die von 1910 bis 2014 geht. Auf diese Weise war es mir möglich zu kontrollieren was der Benutzer einträgt und sicher zu stellen, dass keine fehlerhaften Jahre eingetragen werden.