

# Germany Covid Visualization

untitled

27/03/2021

## Contents

<b>Introduction</b>	<b>2</b>
<b>Data Structure</b>	<b>2</b>
<b>Data Preprocessing</b>	<b>2</b>
<b>Exploratory Data Analysis</b>	<b>3</b>
A. Total cases per state . . . . .	3
B. Cases per Age Group including Gender . . . . .	5
C. Cases per State in Comparison with the Average . . . . .	6

## Introduction

The 2019 Novel Coronavirus (COVID-19) continues to spread in countries around the world. This dataset provides a daily updated number of reported cases & deaths in Germany on the federal state (Bundesland) and county (Landkreis/Stadtkreis) level. The data collected starts from 7th January 2020 to 25th March 2021. The original data are being collected by Germany's Robert Koch Institute and can be download through the National Platform for Geographic Data.

## Data Structure

In this data set, we have a mix of character and numeric variables. It has 8 variables and 653,911 observations.

```
## 'data.frame':   653911 obs. of  8 variables:
## $ state      : chr  "Baden-Wuerttemberg" "Baden-Wuerttemberg" "Baden-Wuerttemberg" "Baden-Wuerttemberg"
## $ county     : chr  "LK Alb-Donau-Kreis" "LK Alb-Donau-Kreis" "LK Alb-Donau-Kreis" "LK Alb-Donau-Kreis"
## $ age_group  : chr  "00-04" "00-04" "00-04" "00-04" ...
## $ gender     : chr  "F" "F" "F" "F" ...
## $ date       : chr  "2020-03-27" "2020-03-28" "2020-04-03" "2020-10-18" ...
## $ cases      : int   1 1 1 1 1 1 3 1 1 1 ...
## $ deaths     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ recovered  : int   1 1 1 1 1 1 3 1 1 1 ...
```

## Data Preprocessing

15,187 missing data points can be found in the age group and gender variable. The volume of missing data points are insignificant and will be removed. The complete data set now has 639,187 observations and 8 variables.

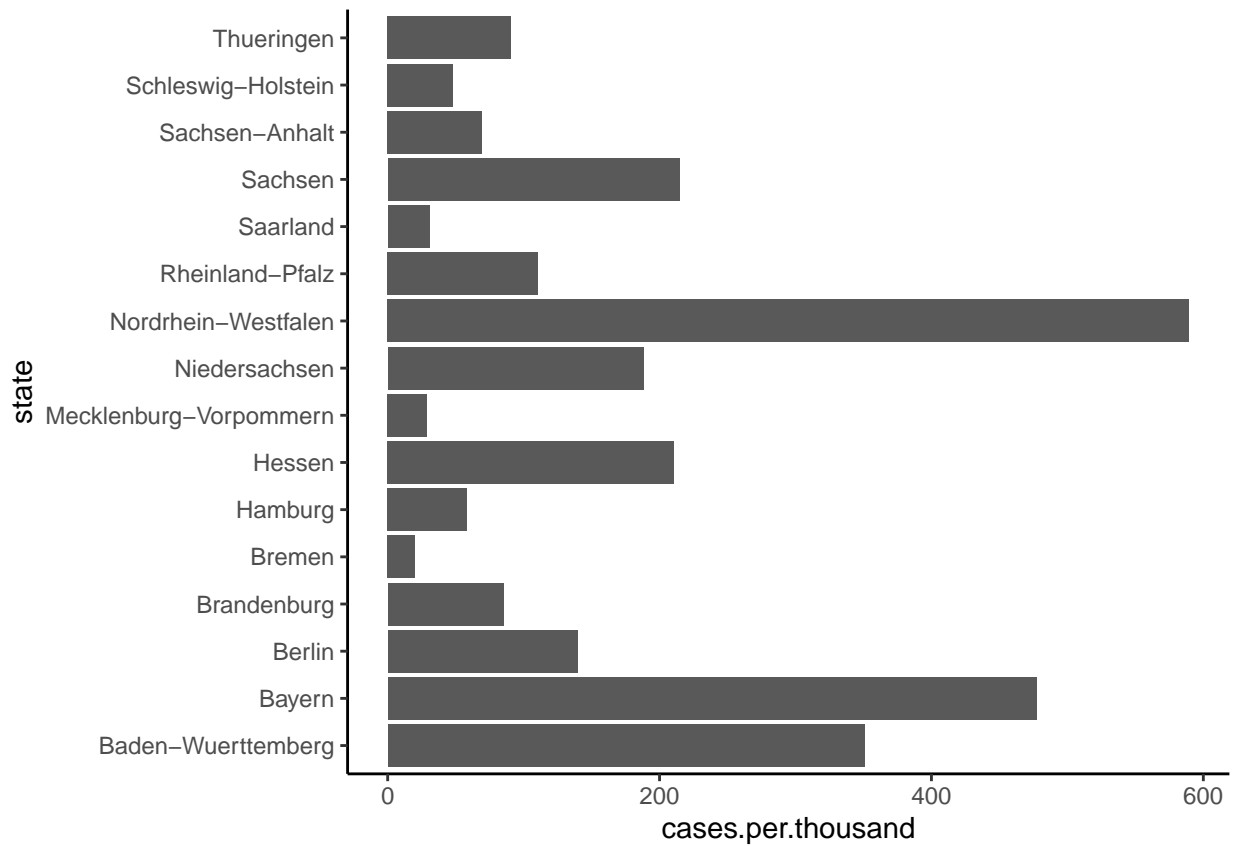
```
## [1] 15187
```

```
##      state    county age_group   gender    date    cases    deaths recovered
##          0          0      2146    13041      0         0         0         0
```

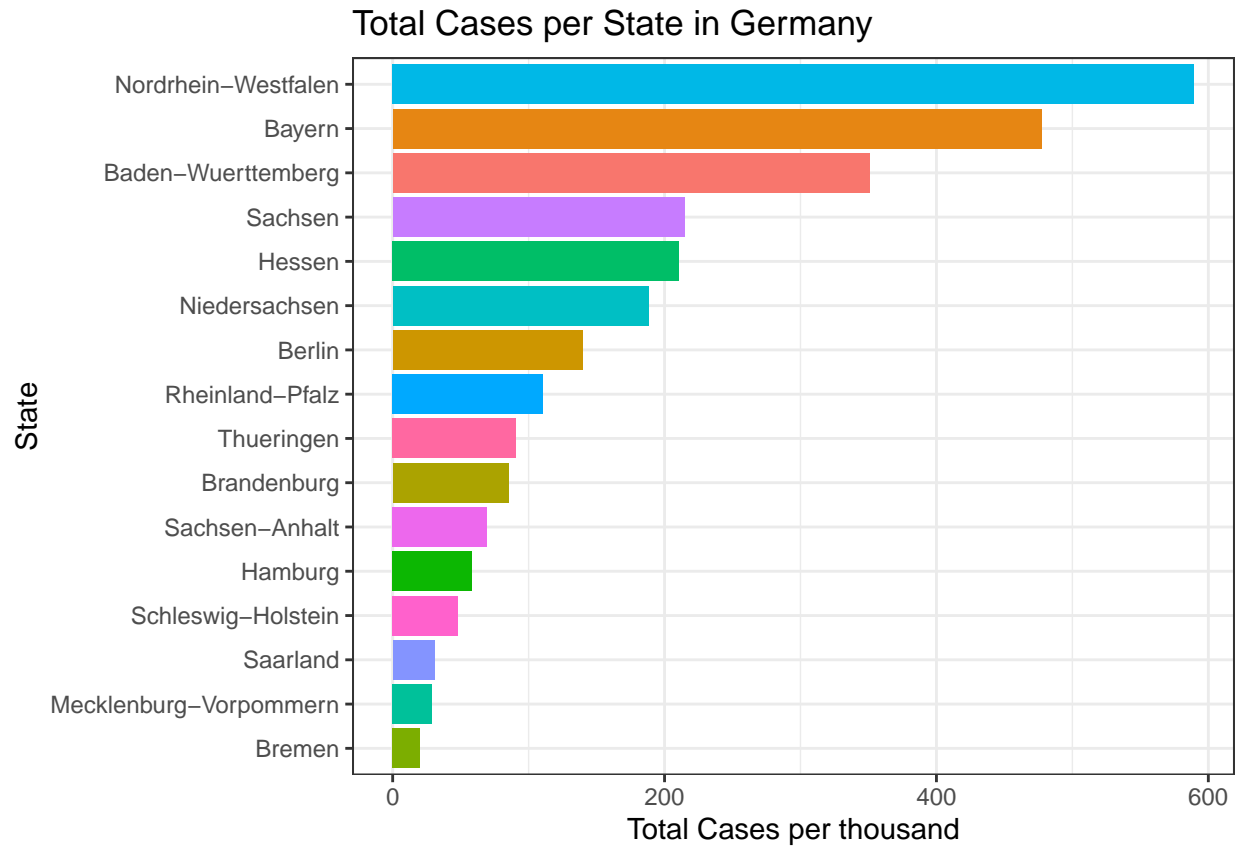
```
## [1] 639187      8
```

# Exploratory Data Analysis

## A. Total cases per state



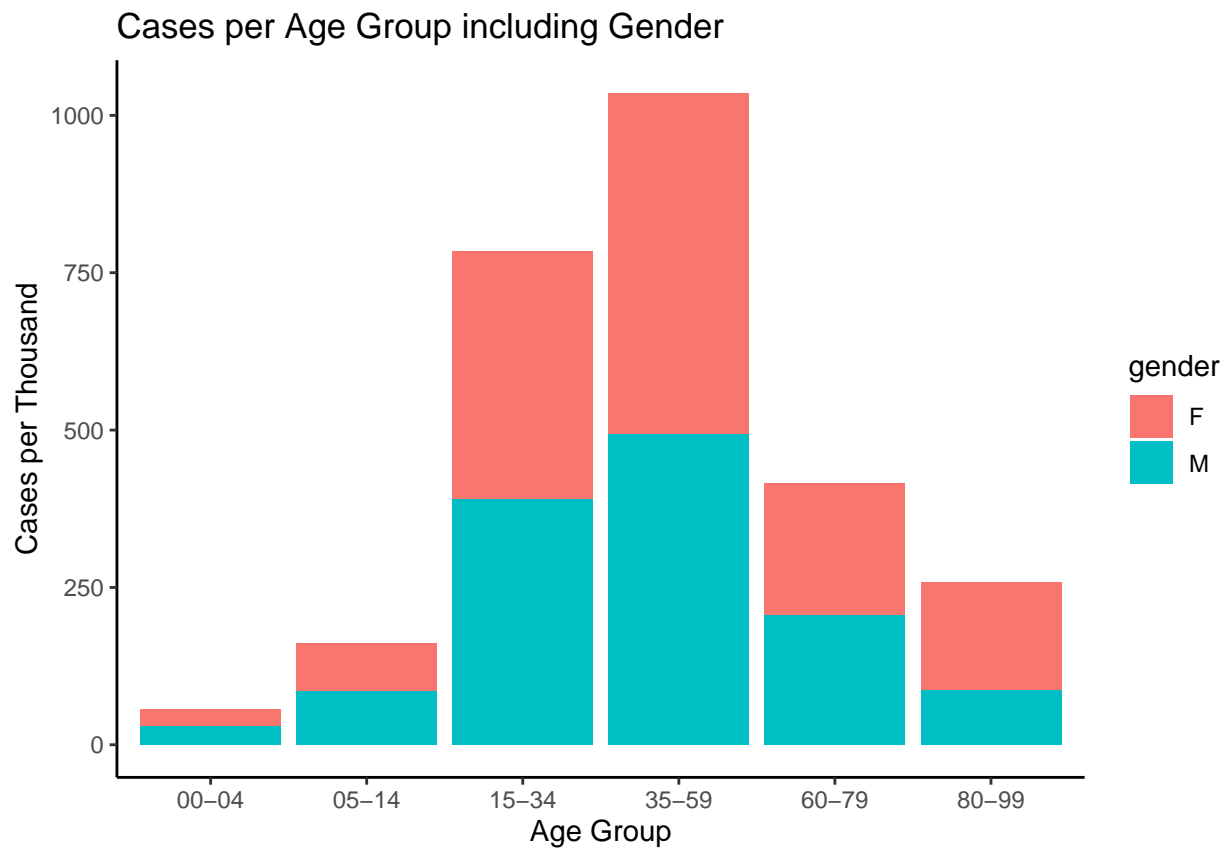
In this diagram, we can see the total amount of covid cases per state in Germany. However, there is no clear order, this will be improved in the next diagram.



This new diagram shows the ordered cases per state in an ordered fashion. Nordrhein-Westfalen is the state with the highest total cases since the inception of the pandemic. Bremen on the other hand has the lowest amount of total cases.

## B. Cases per Age Group including Gender

Now we will visualize the cases per age group accounting for gender distribution in each age group. The age group with the highest total cases is the 35-59 age group, with infections split almost equally amount the sexes.



## C. Cases per State in Comparison with the Average

By normalizing the cases per state we can visualize which states fall above and below the national average for covid cases.

```
## [1] "2020-01-07"
```

```
## [1] "2021-03-25"
```

