

# DM & Visualization - Exam 2023-02-07 - Solution

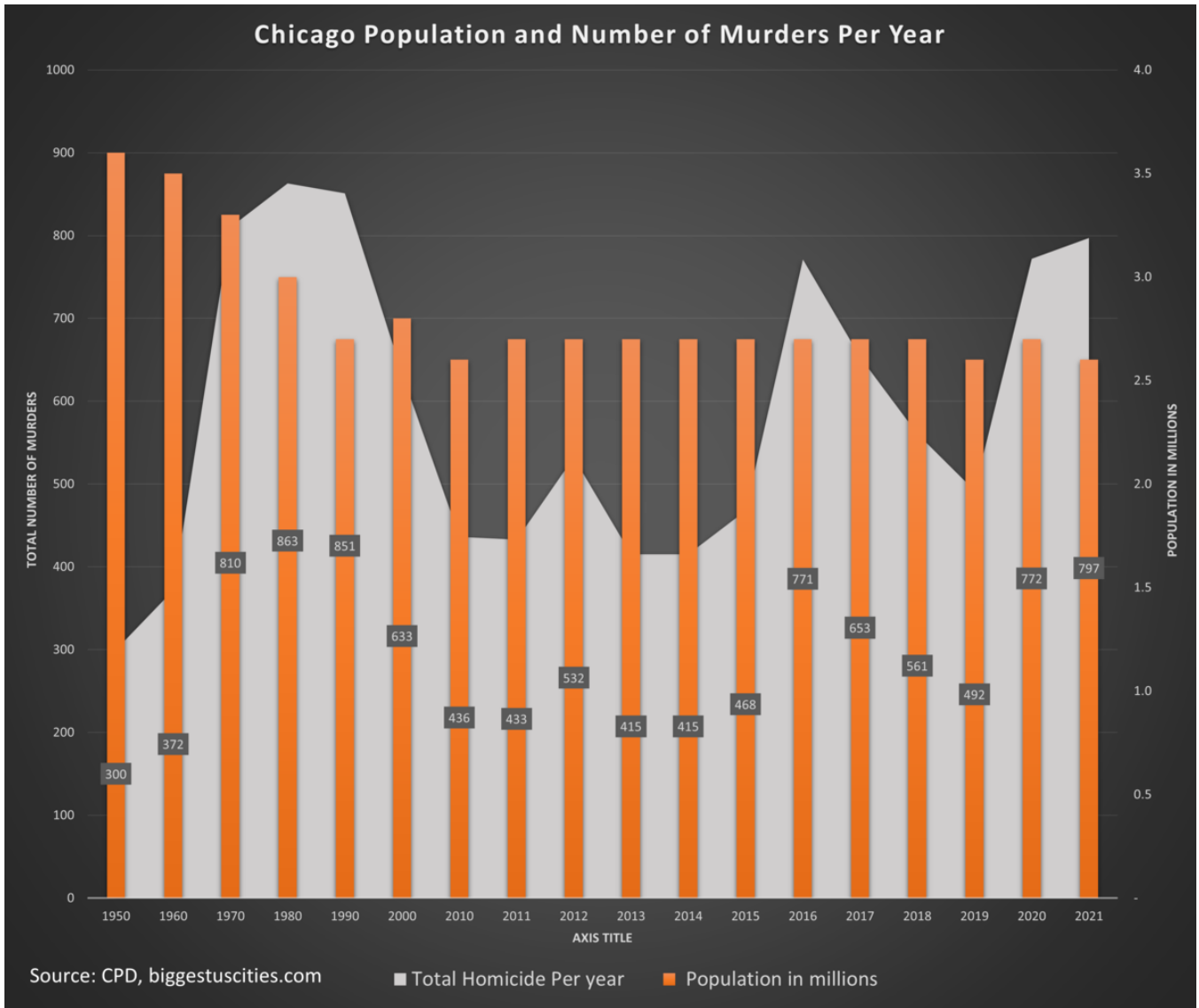


Figure 1: Chicago Population and Number of Murders Per Year

## Analysis

Analyze the above graph according to the following criteria.

### Question

Which one of the following questions represents the purpose of this visualization?

- What is the trend of the Chicago population over several years?
- What is the number of murders per year in Chicago?
- What is the year with the highest number of murders in Chicago?
- *What is the correlation between the number of murders and the population of Chicago over several years?*
- What is the average number of murders in Chicago in the last decade?

### Data

Is the data quality appropriate? Select true answers only.

- The number of murders and the population are accurate data points because they are represented as percentages.
- *Data is not complete because several years are missing.*
- Data is consistent because some decades are compared with single years.
- *Data is reasonably updated for the task because the last data point is from 2021.*
- Data is not credible because no source is reported.
- The visualization is very clear because the labels are associated with the bars.

- Precision is appropriate for the task as the population is reported in millions with a decimal digit.
- Data is complete because starting from 2010 all years are reported.
- Data about murders is credible because the source is the Chicago Police Department (CPD).
- Precision is not appropriate as murders should be represented with more decimal digits.

## Visual

### Proportionality

Are the values encoded in a uniformly proportional way?

The values encoded in the graph are proportional as both y-axis start from zero. Despite the uniform encoding of data, it is challenging to compare measures due to the different scales used. The years displayed on the x-axis are partly decades and partly individual years.

### Utility

All the elements in the graph convey useful information?

The current background color is too dark, hindering the visibility of the numbers on the axis. A lighter, uniform background color would be more suitable. The shading on the bars and the colored area under the line are not necessary and can be removed. The axis title "AXIS TITLE" should be changed to "YEARS" for a more meaningful representation of the data.

### Clarity

Are the data in the graph clearly identifiable and understandable (properly described)?

The use of a double scale and the lack of clear connections between the graph and its scale make it difficult to understand the data. Additionally, the labels on the bars are confusing and do not correspond to the actual values of the bars. The X-axis also lacks a clear title. The current representation presents challenges for color-blind individuals, as the legend is necessary to distinguish the different meanings of the values.

## Design

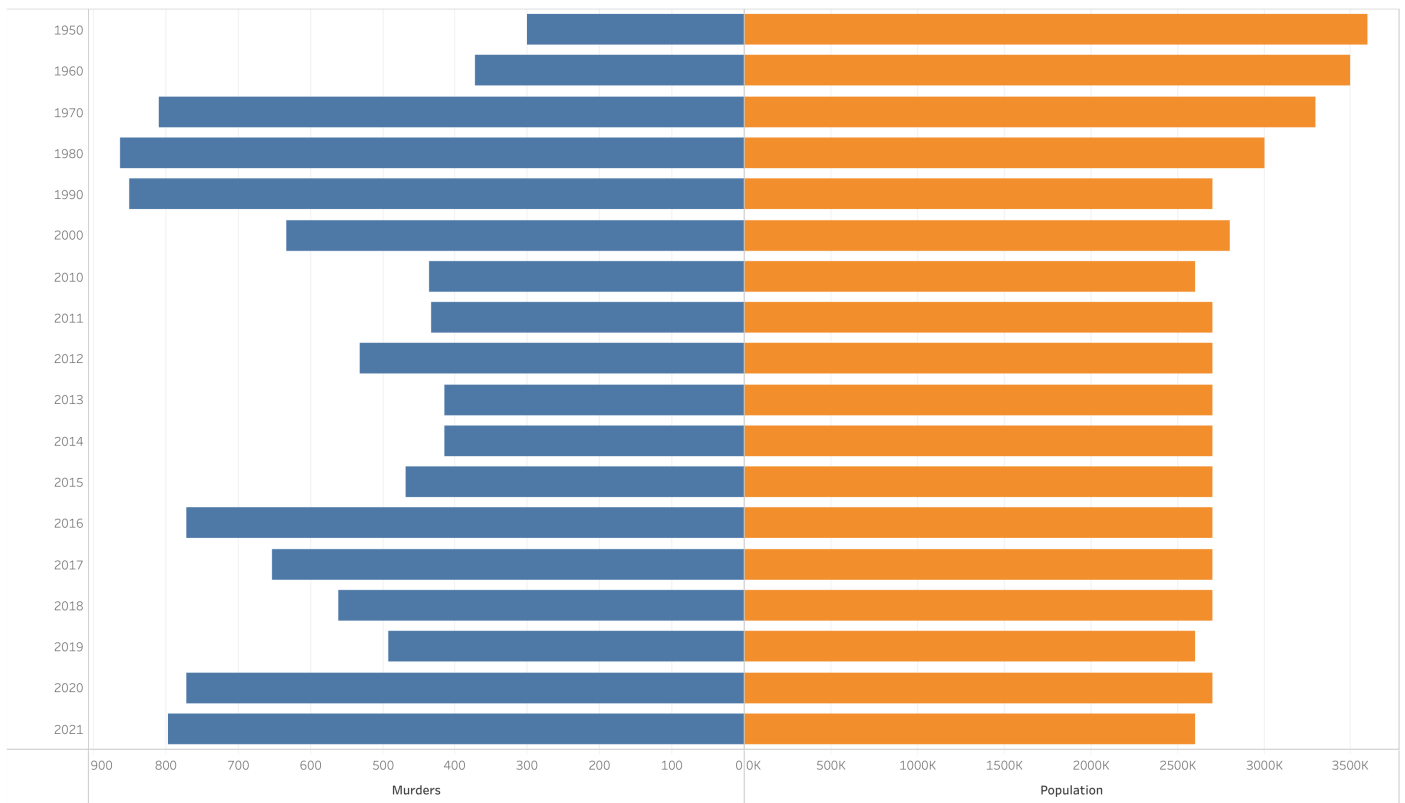
Design the visualization based on the following data structure.

Field	Dim./Measure
YEAR	Dimension
POPULATION	Measure
MURDERS	Measure

## Design schema

Schema	Details
Columns:	SUM(MURDERS), SUM(POPULATION)
Rows:	YEAR(YEAR)
Graph type:	Bar
Color:	Measure Names
Size:	Default
Label:	Default

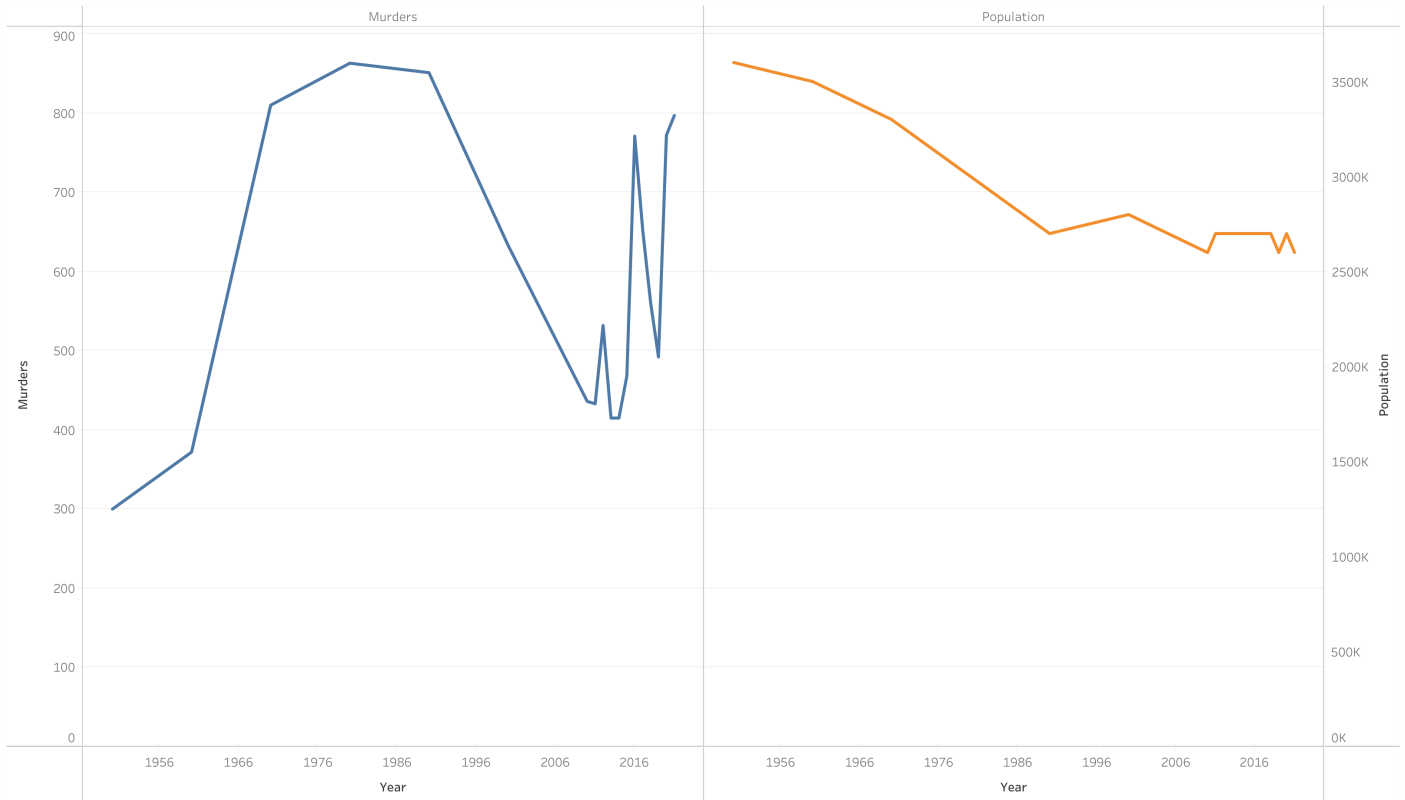
## Sketch of the resulting graph



## Design schema

Schema	Details
Columns:	Measure Names, YEAR(YEAR)
Rows:	SUM(MURDERS), SUM(POPULATION)
Graph type:	Line
Color:	Measure Names
Size:	Default
Label:	Default

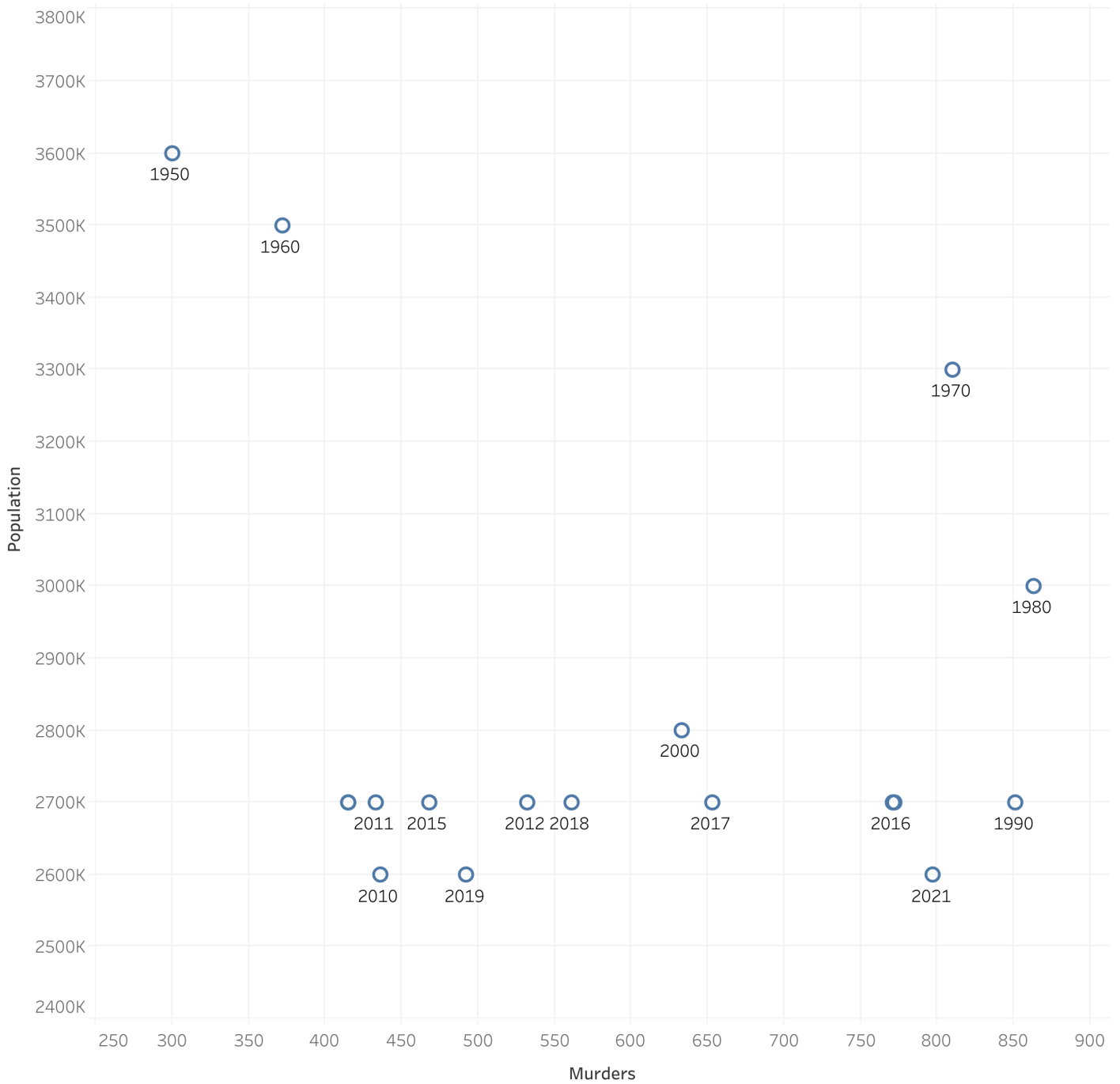
## Sketch of the resulting graph



## Design schema

Schema	Details
Columns:	SUM(MURDERS)
Rows:	SUM(POPULATION)
Graph type:	Shape
Color:	Default
Size:	Default
Label:	YEAR(YEAR)

## Sketch of the resulting graph



## Theory

Which one of the following elements should not be removed from a visualization according to the guidelines for visual utility?

- Heavy grid lines
- Logos and other pictures
- Decorative colors
- *Separated legend*
- Background shades