Gender and Age Group Analysis of Mortality Patterns in Alberta
Group 4

Team members:
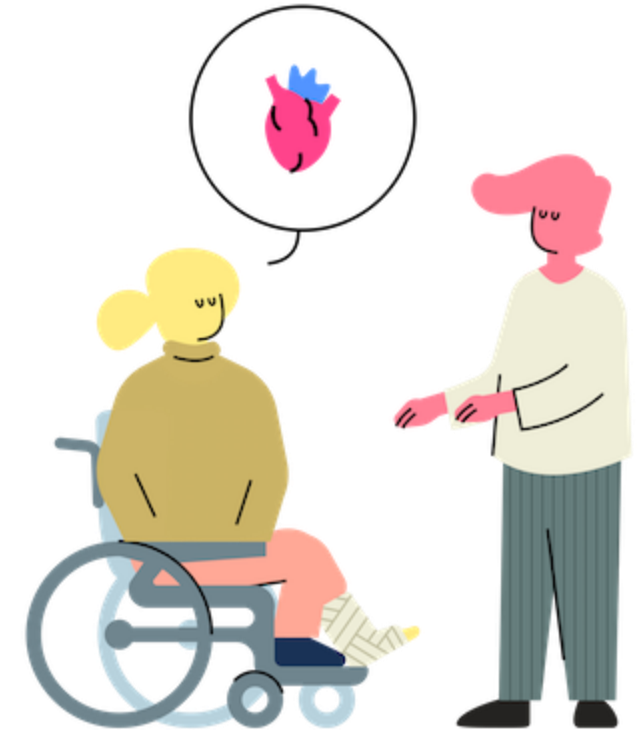Osama Bilgrami
Graeme Ko
Umair Qureshi
Soyebal Saad Adnan
Ray Alsaidi

Introduction

- Our Company: Cradle2Grave Analytics
- Thriving Alberta
- Government of Alberta
- Premier Koala-fied.

Guiding Questions

- What are the main causes of death?
  - By gender?
  - By year? Overall (2001-2022)?
  - By age group?

- Is there a trend?

- What can the data tell us about the future?

Dataset

- Dataset via Alberta's Open Government program.

- We have the right to use and manipulate the dataset provided we give proper citation, as seen in the references section.

Column Title	Data Type	Description
Calendar Year	Categorical	The year that the death occurred (2001-2021)
Cause	Categorical	Causes of death (538 Unique cases)
Gender	Categorical	Represents the gender of the person (F,M,T,U)
Age Groups	Categorical	Age group of the person (ranges from 0-90+)
NS	Numeric	Number of casualties with unknown age group
Res Total	Numeric	Number of casualties who are a resident of Alberta
Non Res Total	Numeric	Number of Casualties who are not a resident of Alberta
Grand Total	Numeric	Total number of Casualties

Data Cleaning

- Data types: int64 and object
- Type-casted 'Calendar Year' to datetime and the other columns to int64
- Checked for missing values
- Created a function to strip serial number from "Causes" column and whitespace from the dataframe

|	Calendar Year	Cause	Gender	0-4	5-9	10-14	15-19	20-24	25-29	30-34
0	2001	1 Cholera	M	0	0	0	0	0	0	0
1	2001	1 Cholera	F	0	0	0	0	0	0	0
2	2001	1 Cholera	T	0	0	0	0	0	0	0
3	2001	2 Typhoid fever	M	0	0	0	0	0	0	0
4	2001	2 Typhoid fever	F	0	0	0	0	0	0	0

Data Cleaning

- Removed 'Total' and 'NS' rows which were double-counting

- Cleaned out mistyped and unknown entries in the 'Gender' column

- Used 'pd.melt' to transform the dataset from a wide format to a long format

	year	cause_death	gender	age_group	count
0	2001-01-01	Cholera	M	0-4	0
1	2001-01-01	Cholera	F	0-4	0
2	2001-01-01	Typhoid fever	M	0-4	0
3	2001-01-01	Typhoid fever	F	0-4	0
4	2001-01-01	Paratyphoid fevers and other salmonella infect...	M	0-4	0
...	...	...	...	...	...
296471	2022-01-01	Volume depletion, disorders of fluid, electrol...	F	90+	10
296472	2022-01-01	Water transport accidents	M	90+	0
296473	2022-01-01	Water transport accidents	F	90+	0
296474	2022-01-01	Zoster (herpes zoster)	M	90+	1
296475	2022-01-01	Zoster (herpes zoster)	F	90+	0

296476 rows × 5 columns
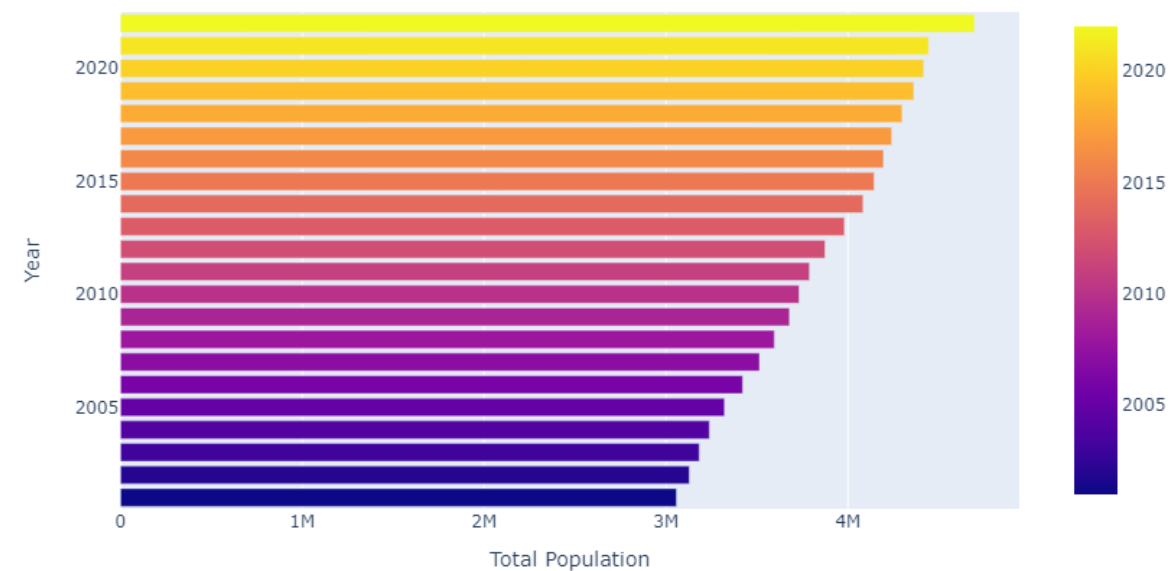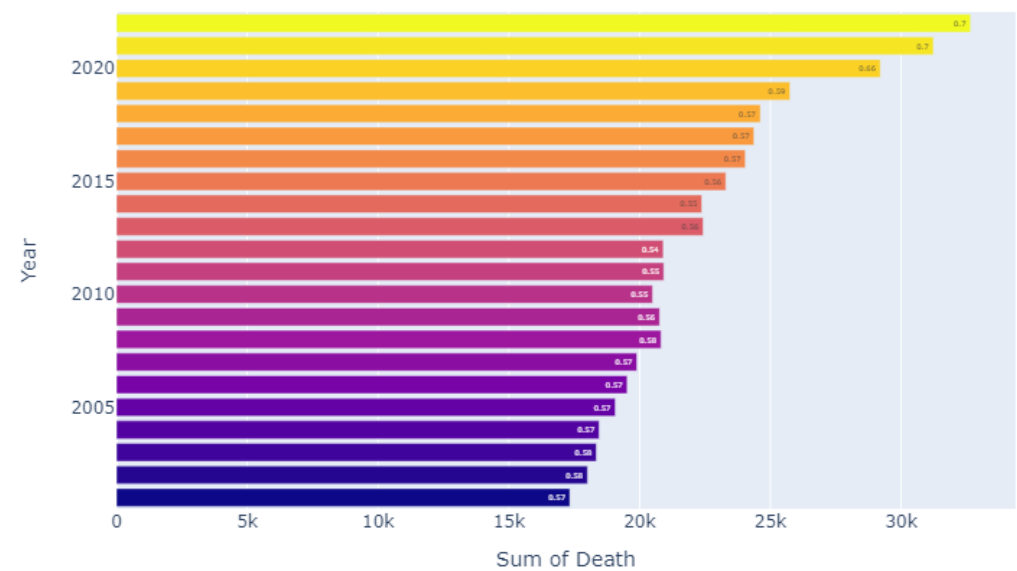
Proportion of deaths
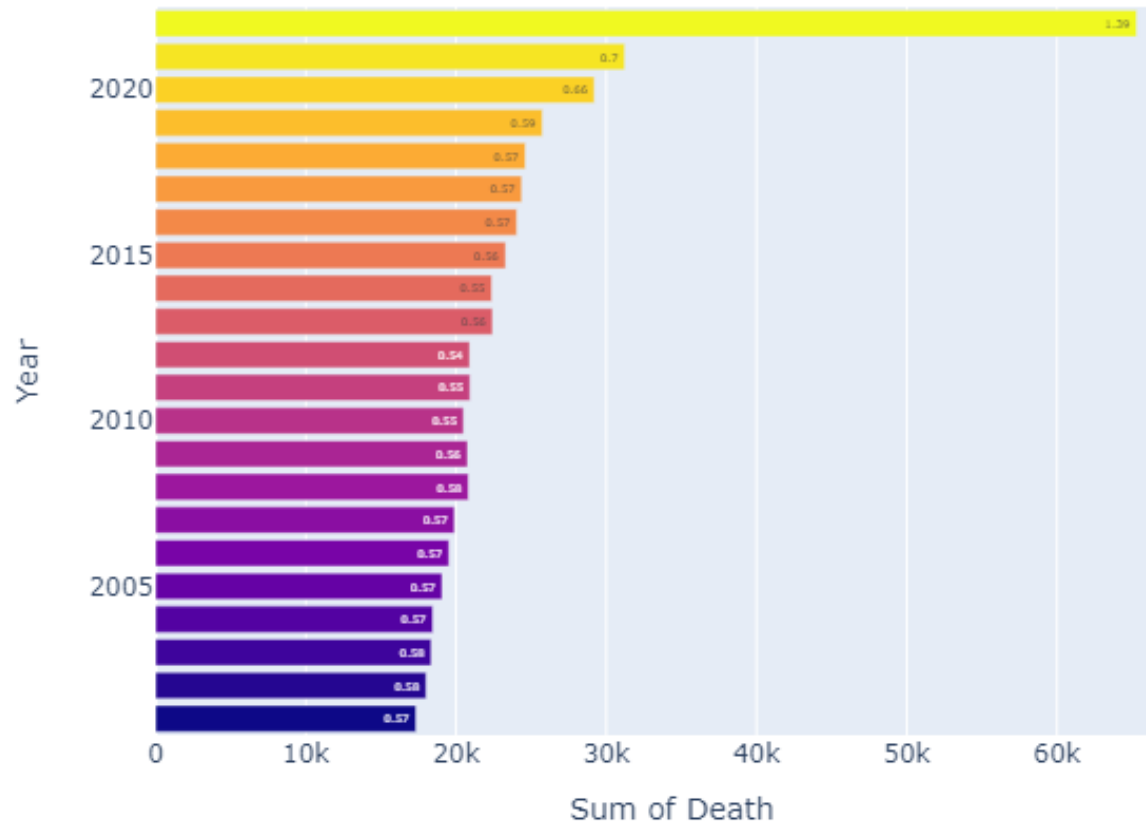
Death rate always has a consistent value of around 0.55%
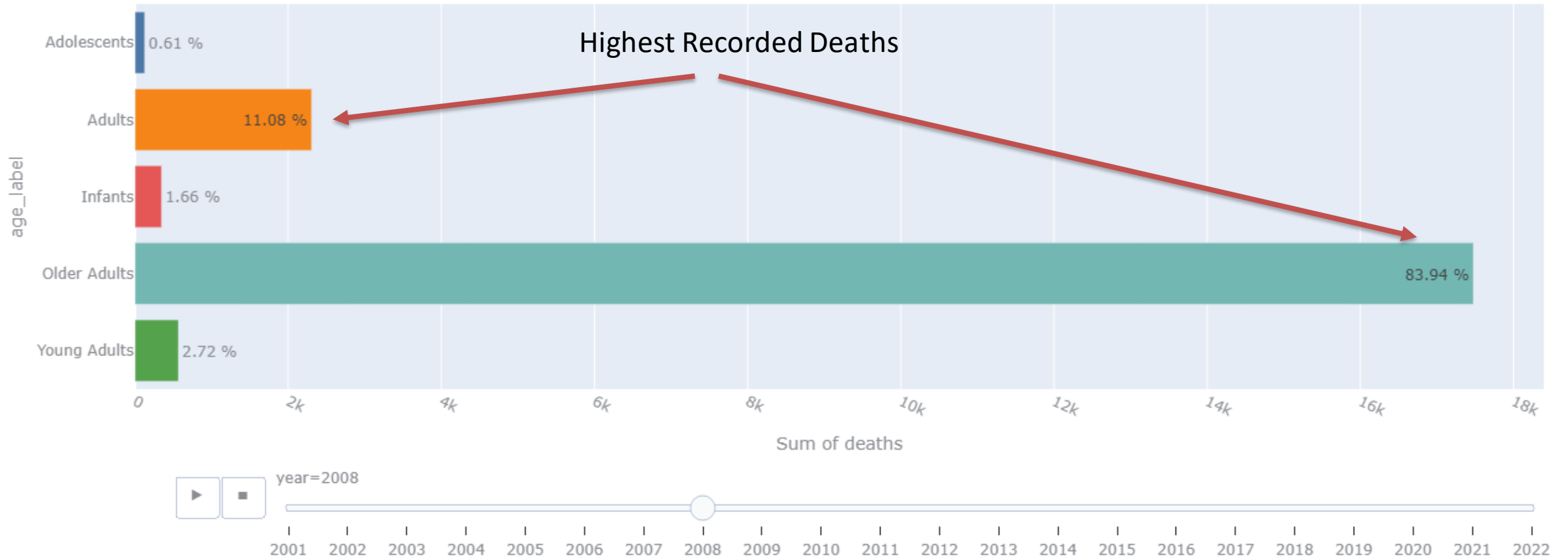
The population increases so does the death rates and so it has a steady value

Proportion of deaths and Total Population from 2001 to 2022

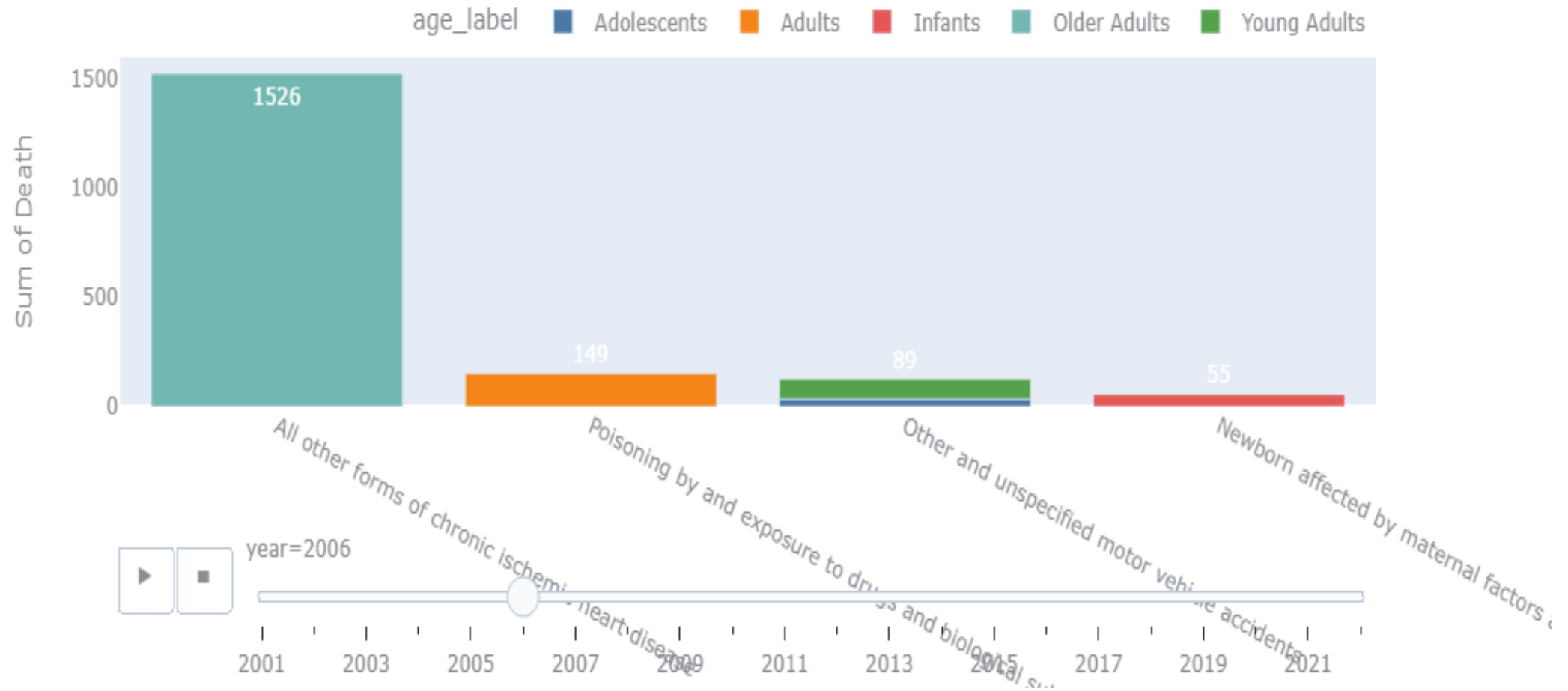Proportion of deaths of each age group from 2001 to 2022

Older adults always has the highest deaths among all age groups - greater than 80% (this number keeps increasing)
2nd Highest number of deaths are from adults but always below 12%
Adolescents has the least deaths recorded – below 1%

Top Causes of Death in each age group

For adults and Young Adults till 2003 the leading cause of death was Suicide which later for most year became poisoning due to drugs and biological substances
The death of adolescents is the least among all categories and the leading cause of their death is motor accidents
The Older Adult group has the most deaths every year with the top cause being Ischemic Heart Disease till 2013 and then it being Organic dementia after that

Trend of Top 3 Death Causes

Males

Females
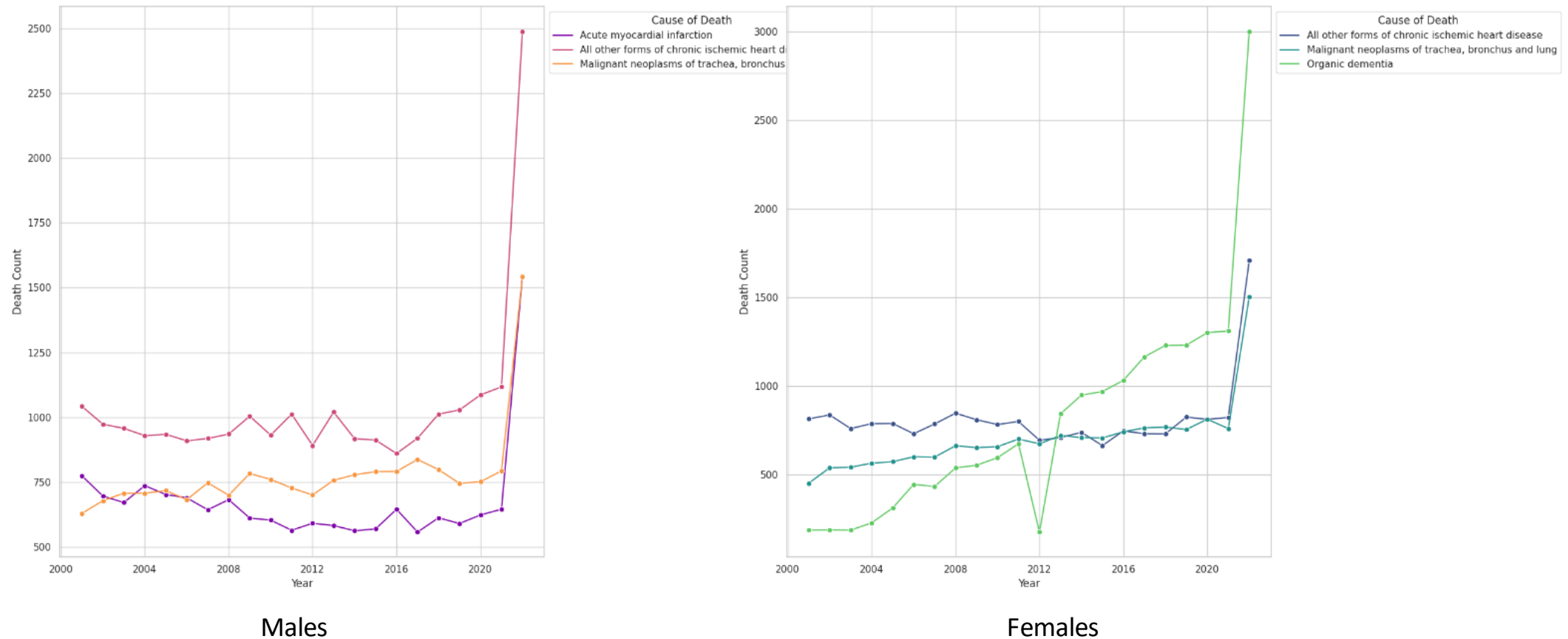
- For both chronic ischemic heart diseases and malignant neoplasms, there is a consistent and stable trend with a slight upward trajectory. This trend closely mirrors the trends observed among males, and the data does not reveal any unexpected patterns.

- For both genders, all three trends exhibited a sharp increase during the global COVID-19 pandemic. This sudden surge was not included in our trend analysis as it is considered outlier data points, but it was displayed in the visualization.

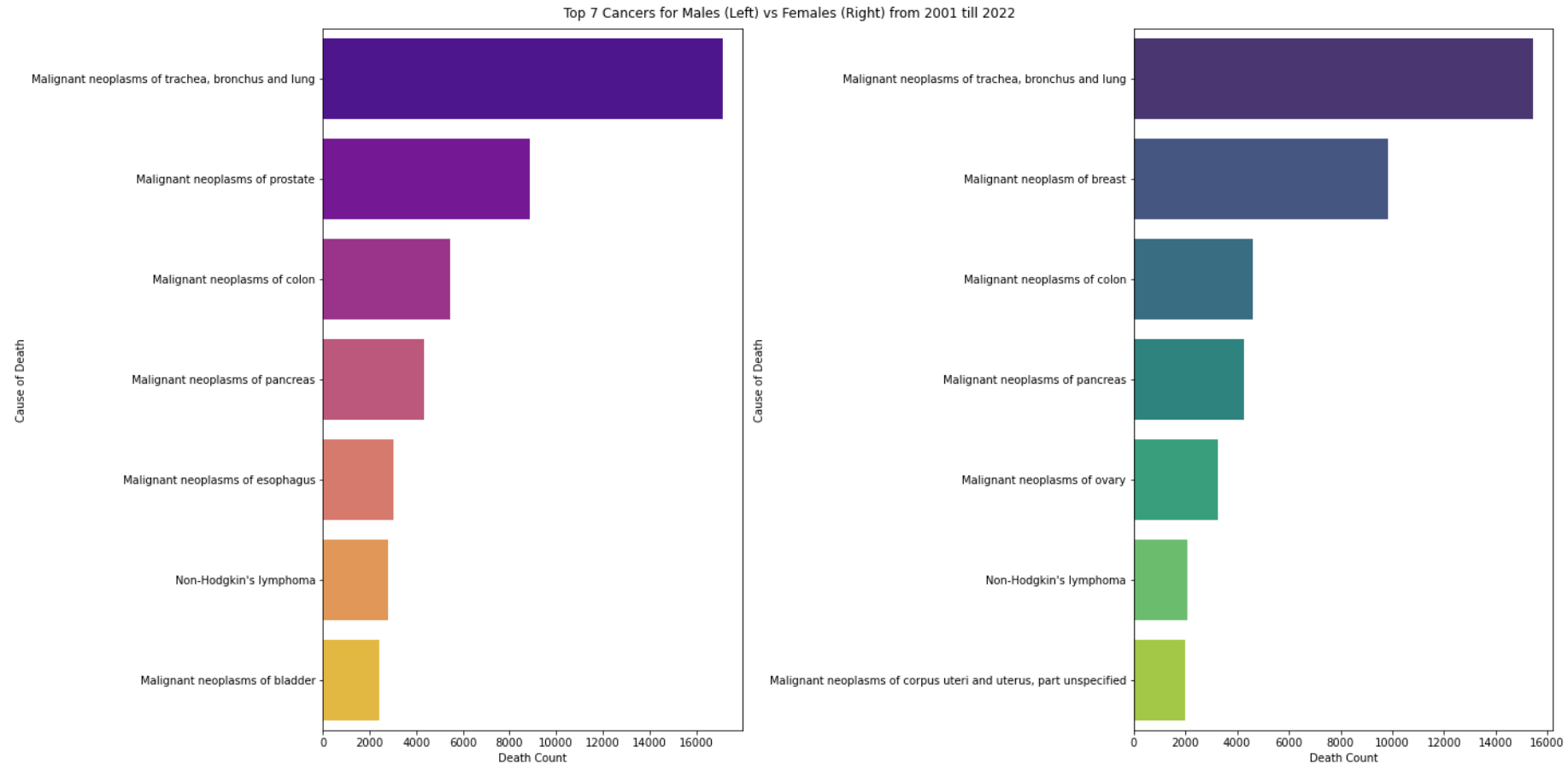# Death by Heart Disease across Age Groups 2001 to 2021
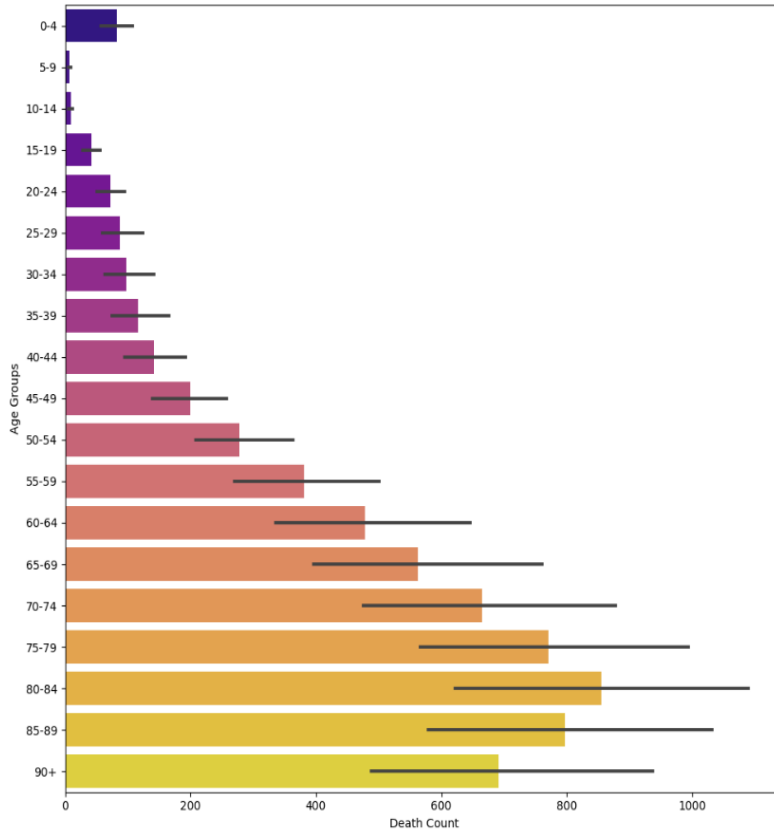
Males
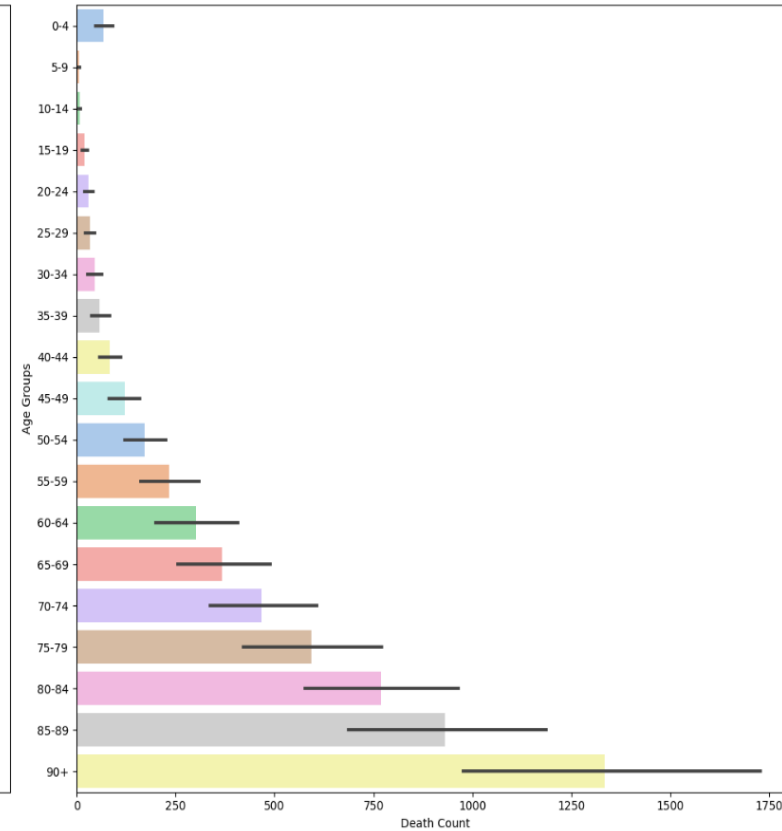
Females

- Largely age related.

- Initial spike with infants (0-4 years).

- A steady increasing trend.

- More men tend to develop heart diseases at a younger age then women.

Cancer: Gender and Age

- The gender difference was found largest in age groups spanning 60-89 years, whereas some age groups closer to being middle-aged showed higher female death counts.

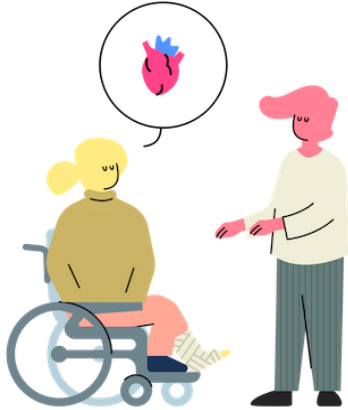- The 75-79 age group appears to be at highest risk.

Cancer-Related Deaths 2001 - 2022 by Age Group
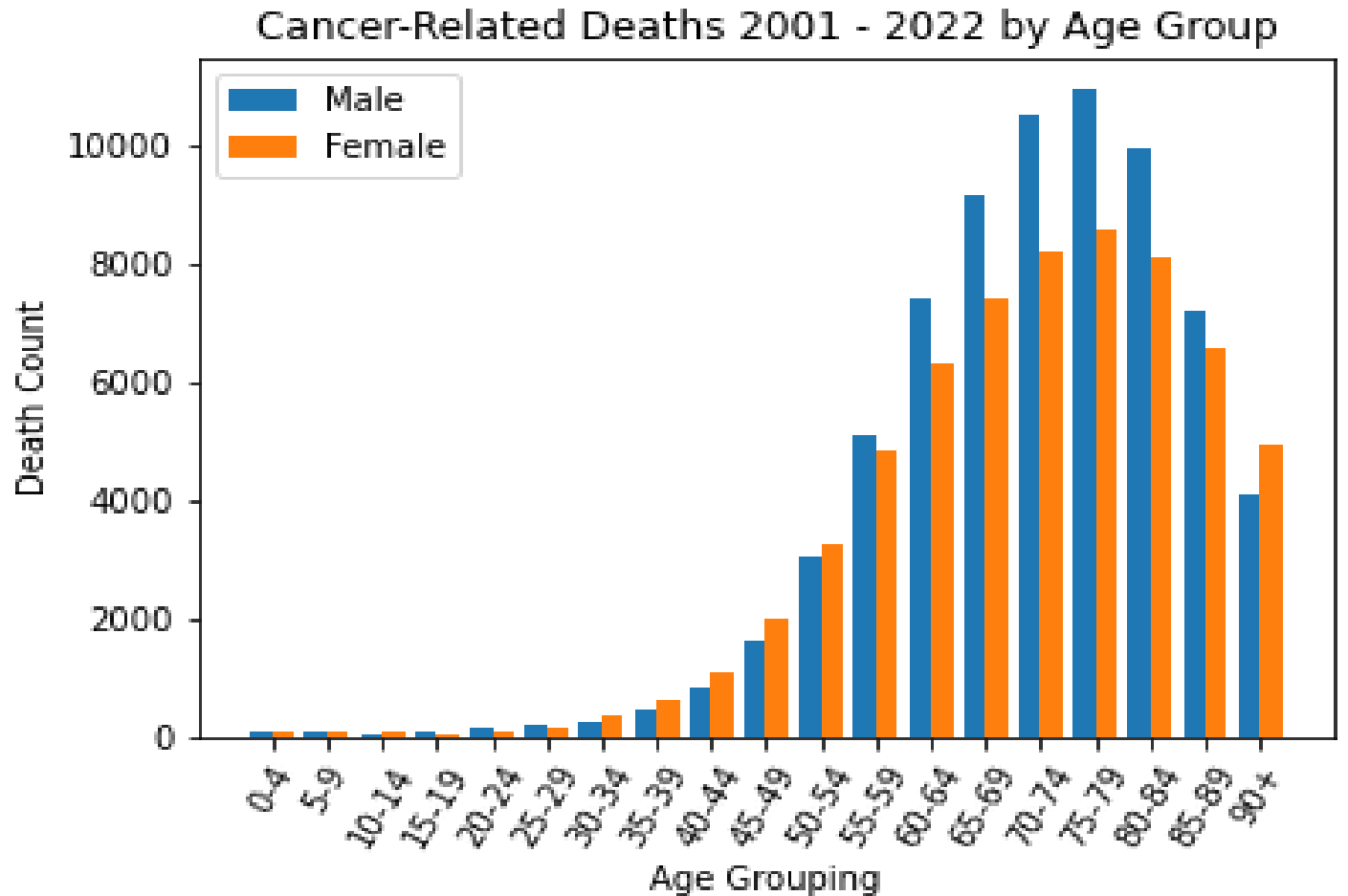
Cancer: The Major Players

Top 7 Cancers for Males (Left) vs Females (Right) from 2001 till 2022

Lung associated cancers had by far the highest mortality rate out of 81 unique cancer related causes we identified.

Conclusions and Next Steps

1 Increase cancer funding

2 Action plan to prevent heart disease

3 Investigate the upward trend of dementia

4 Suicide intervention

5 Drug prevention initiatives

Questions

1. **What does the colors in the proportion of death graph represent?**

- Colors in the graph represent different years for visual appeal

2. **Details on that graph and how did we get the 0.55%?**

- The graph and percentage labels on each bar show the consistency
- The graph compares the number of recorded deaths in each year to Alberta's total population
- The consistent value of around 0.55% represents the proportion of deaths to the total population (ratio)
- This ratio is generally stable, except for 2021-2022 (likely due to COVID)The trend suggests that as the population increases, the number of deaths also rises while maintaining the steady rate

3. **What is causing the spike in 2022?**

4. **How do we group cancer causes together?**

- Keyword matching
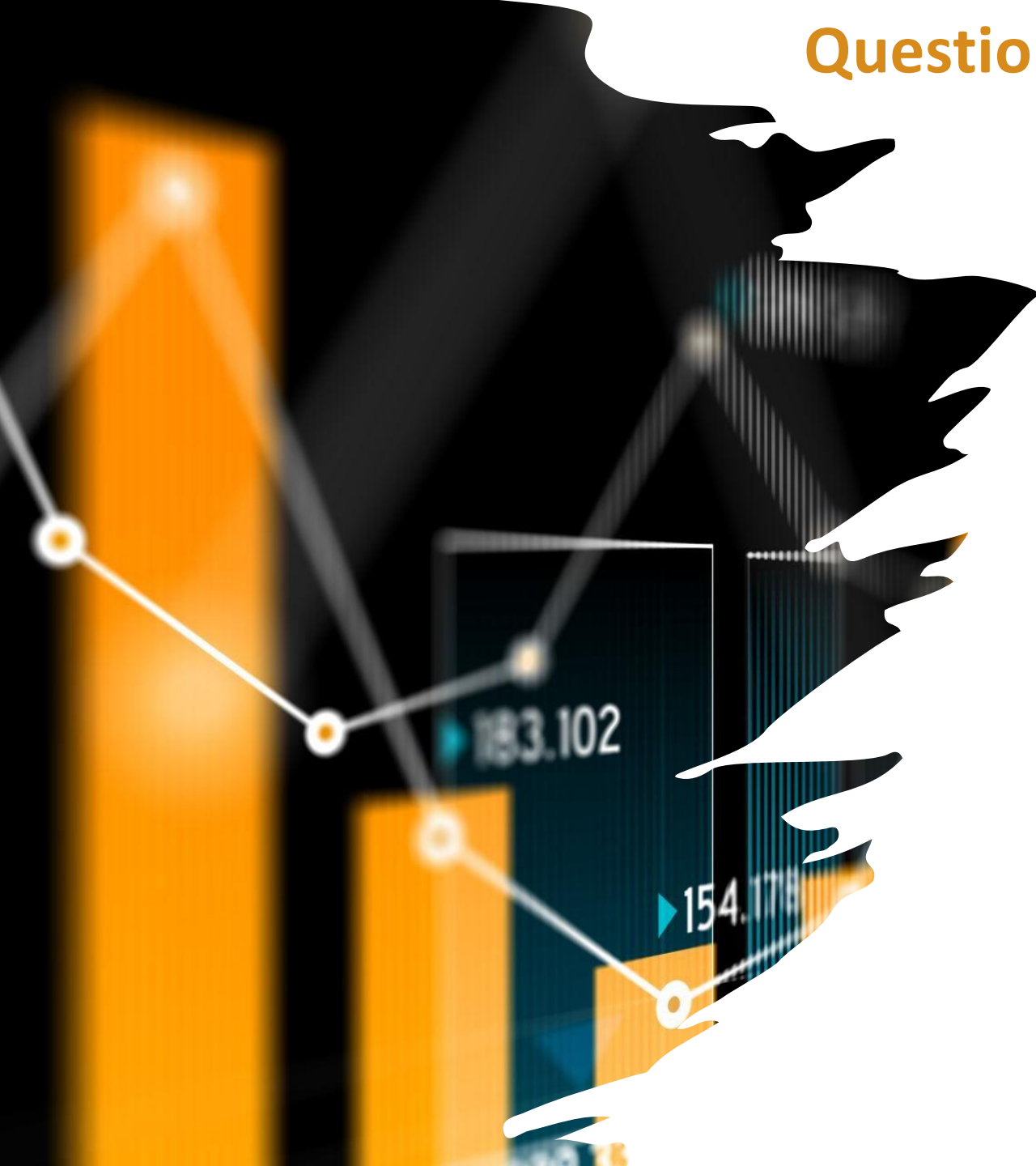
Thank You

In this world nothing can be said to be certain, except death and taxes.

Franklin, in a letter to Jean-Baptiste Le Roy, 1789[1]