

# DATA VISUALIZATIONS USING IBM COGNOS ANALYTICS

Monal Gosai

# WHY USE DATA VISUALIZATIONS

- ▶ Comprehend information quickly
- ▶ Identify relationship and patterns
- ▶ Pinpoint emerging trends
- ▶ Results communicates itself

# SELECT VISUALIZATION TYPE AND PLOT COLUMNS

**Visualizations**

The visualization selection interface includes:

- Navigation paths:** Who Nrevss Combined Prior To 2015 16
  - Region
  - Year
  - Week
  - Total Specimens
  - Percent Positive
  - A ( 2009 H 1 N 1 )
  - A ( H 1 )
  - A ( H 3 )
  - A ( Subtyping Not Performed )
  - A ( Unable To Subtype )
  - B
  - H 3 N 2v
- x-axis:** Length \*
- # Line position \***
- Repeat (column)**
- Repeat (row)**
- Local filters**

**Line and column**

**Build your visualization**  
Drag and drop columns here or onto the slots.

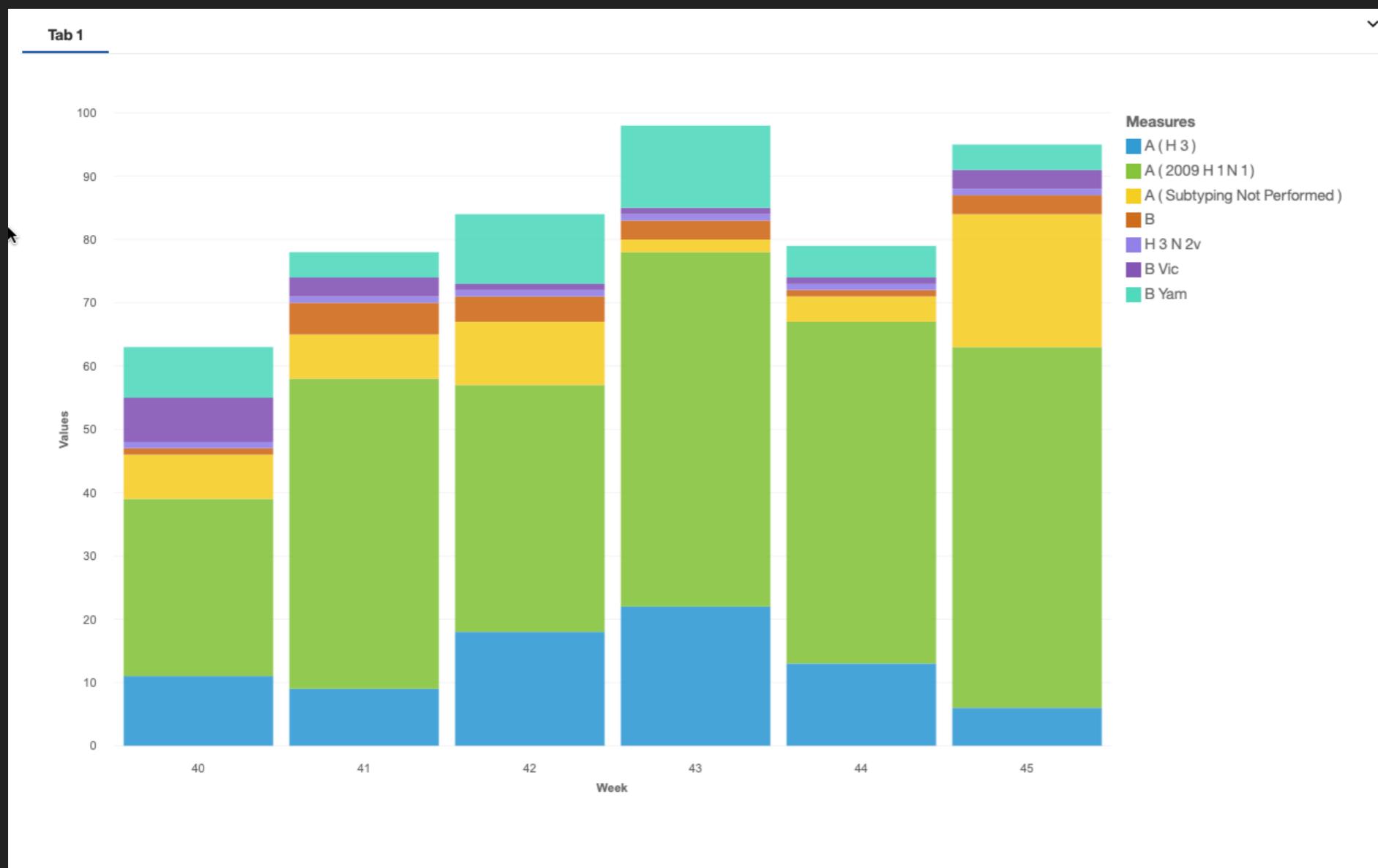
## DATA SET USED

### FLUVIEW INTERACTIVE SURVEILLANCE DATA

- ▶ ILI AND VIRAL
- ▶ HOSPITALIZATION
- ▶ ILI ACTIVITY MAP
- ▶ PAEDIATRIC MORTALITY
- ▶ P&I MORTALITY
- ▶ NOVEL FLU A
- ▶ GEOGRAPHIC SPREAD

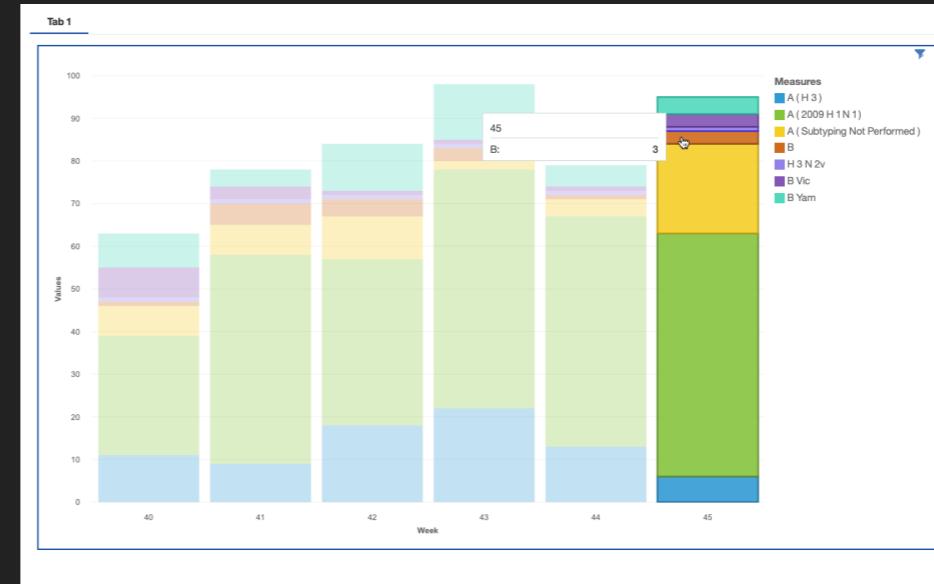
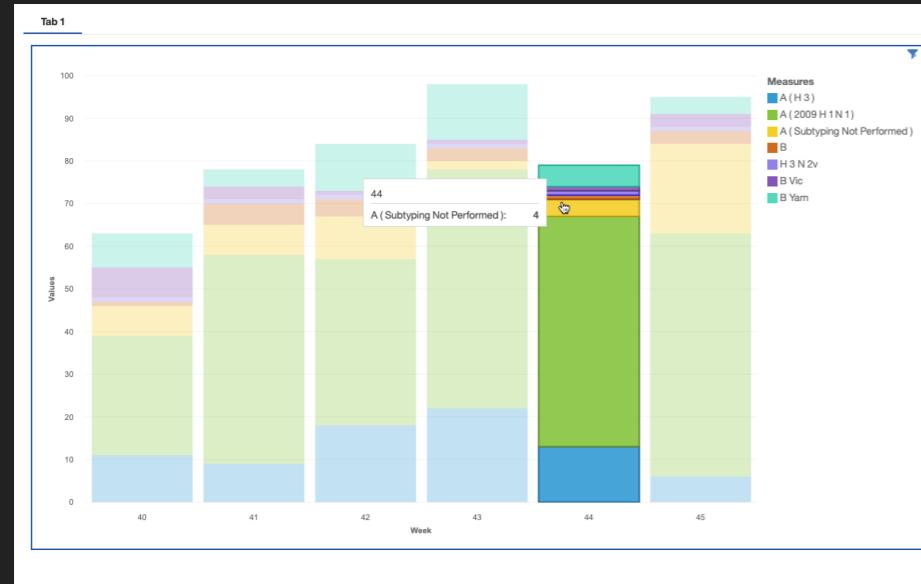
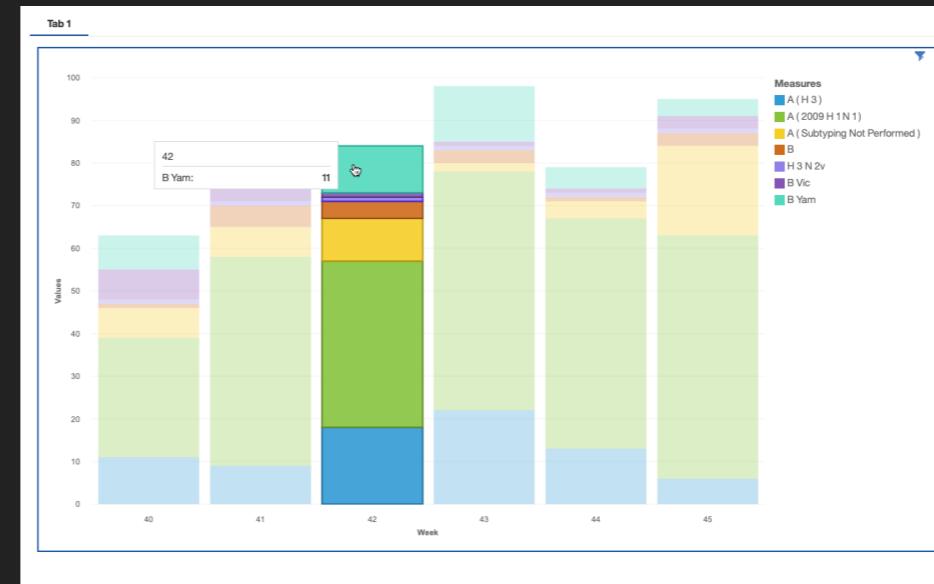
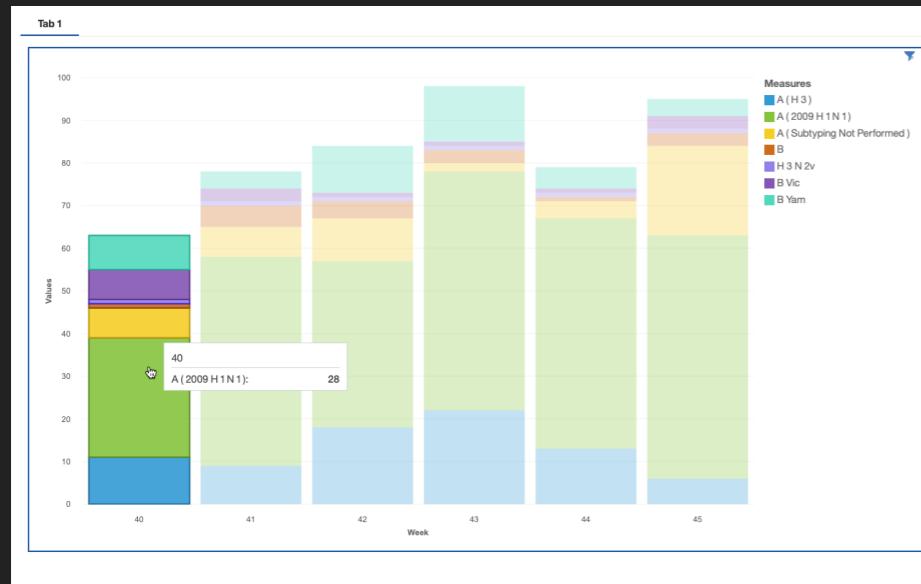
# ILI AND VIRAL

Influenza Positive Tests Reported to CDC by Public Health Laboratories, National Summary, 2018-19 Season, week ending Nov 10, 2018.



Here each color represents different entity

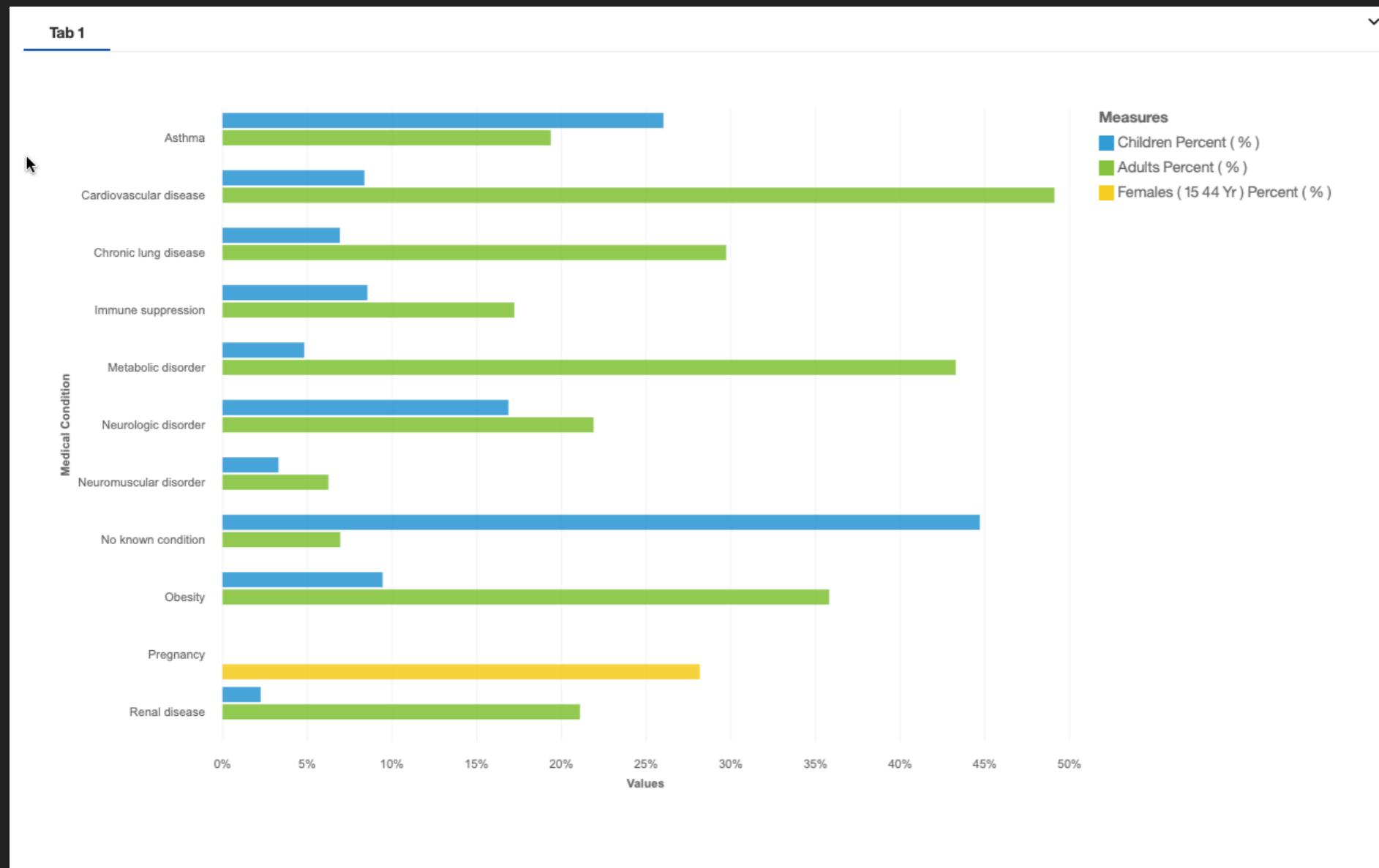
# CUSTOMIZED VIEW OF EACH BARS



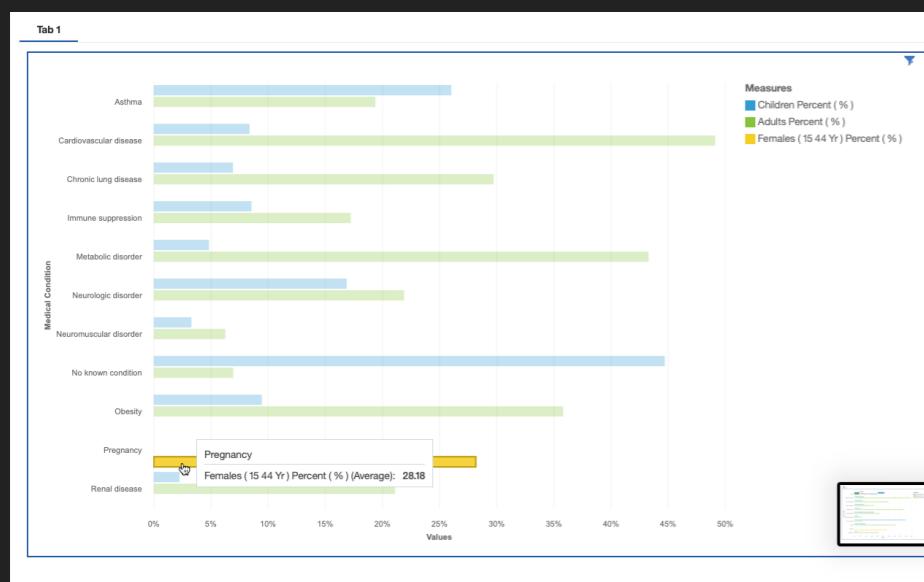
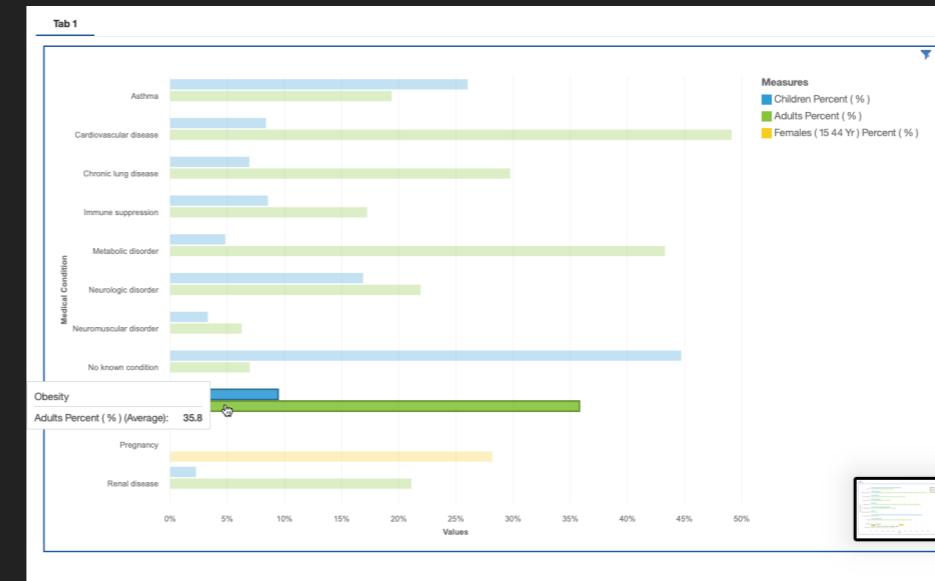
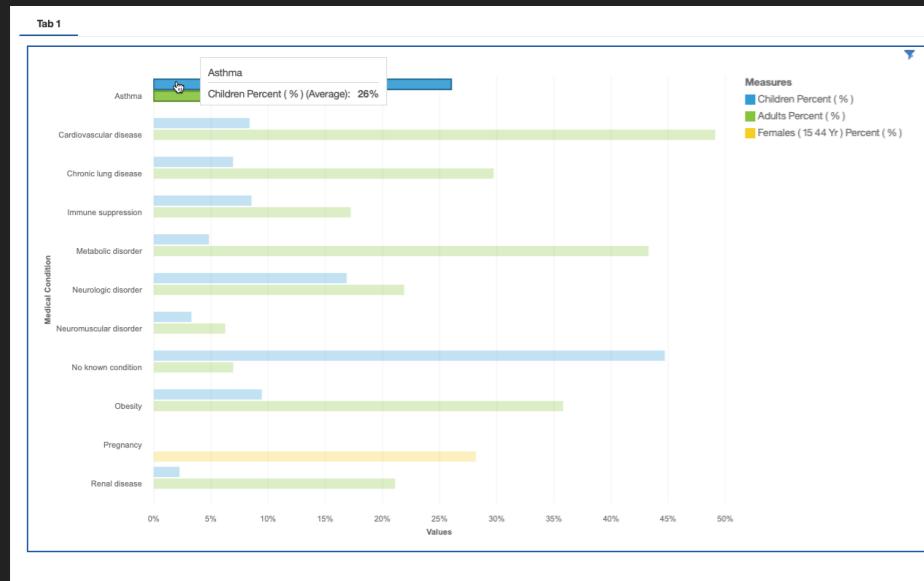
Hovering the cursor in each color reveals detailed information about the entity.

# HOSPITALIZATION

Influenza-Associated Hospitalizations surveillance with horizontal Bar graph.



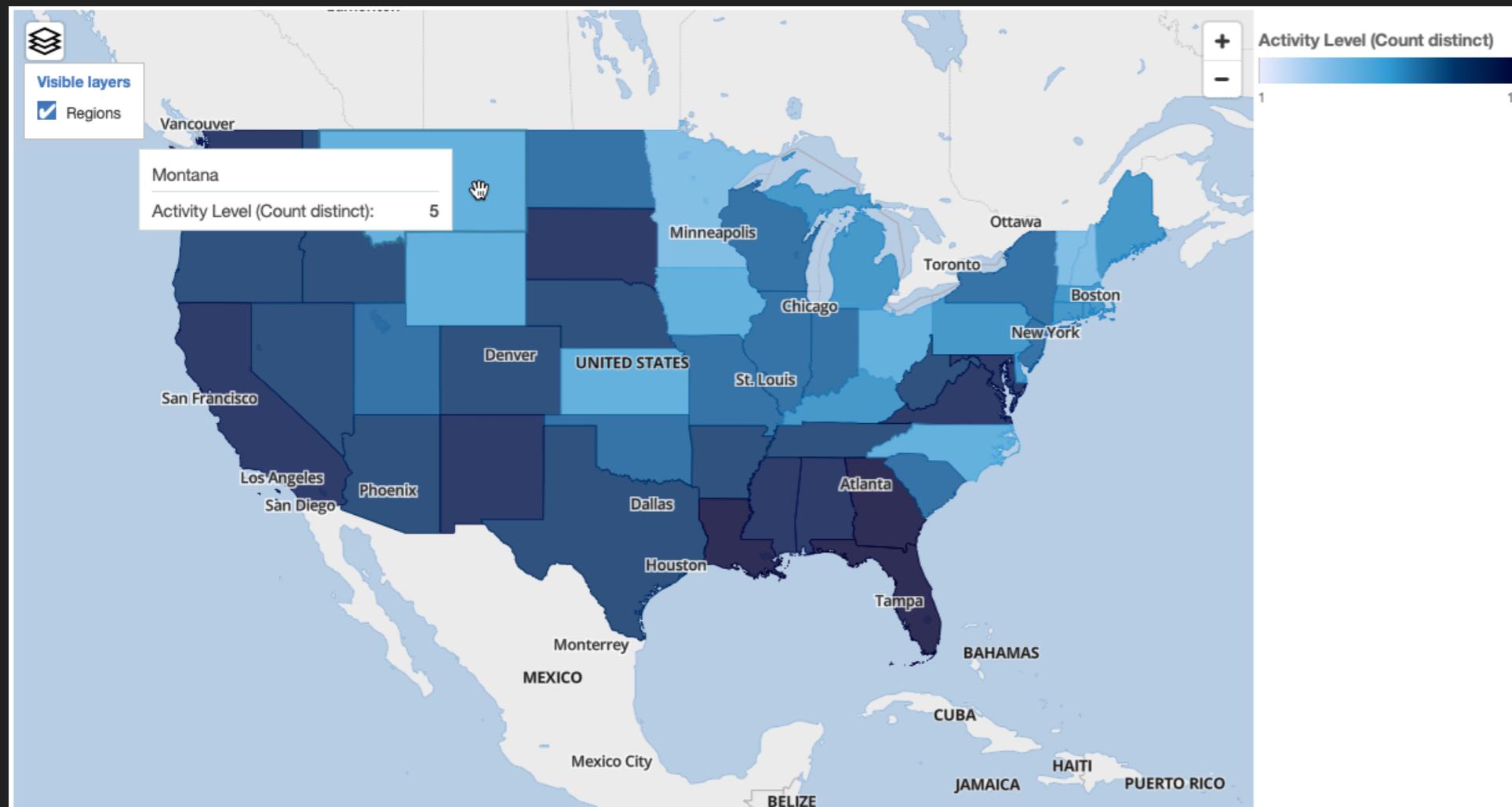
# HOVERING REVEALS DETAILS



Each color reveals different gender and children and during what disease they were hospitalized.

# INFLUENZA-LIKE ILLNESS (ILI) ACTIVITY MAP

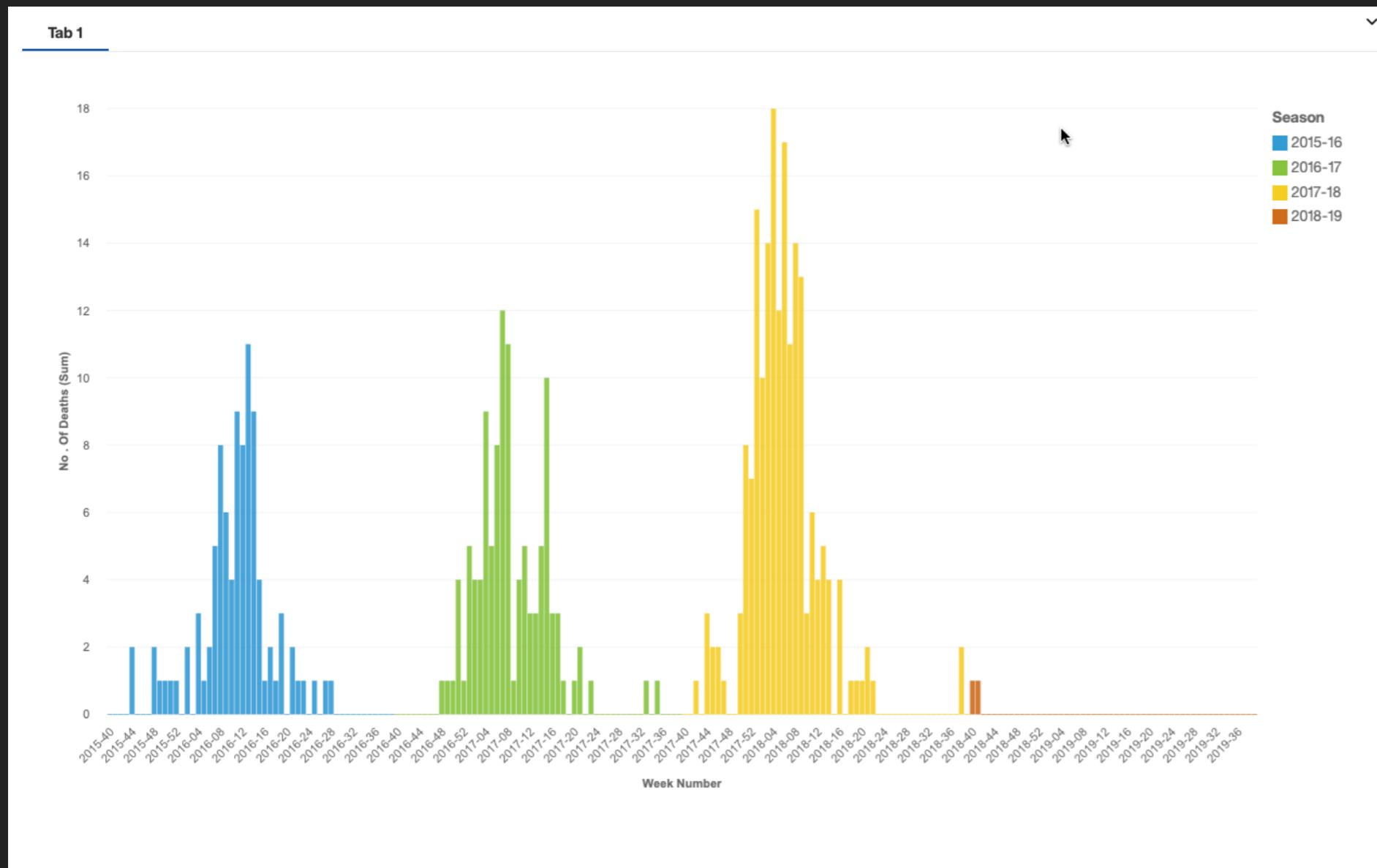
Activities Related to ILI represented in a Map



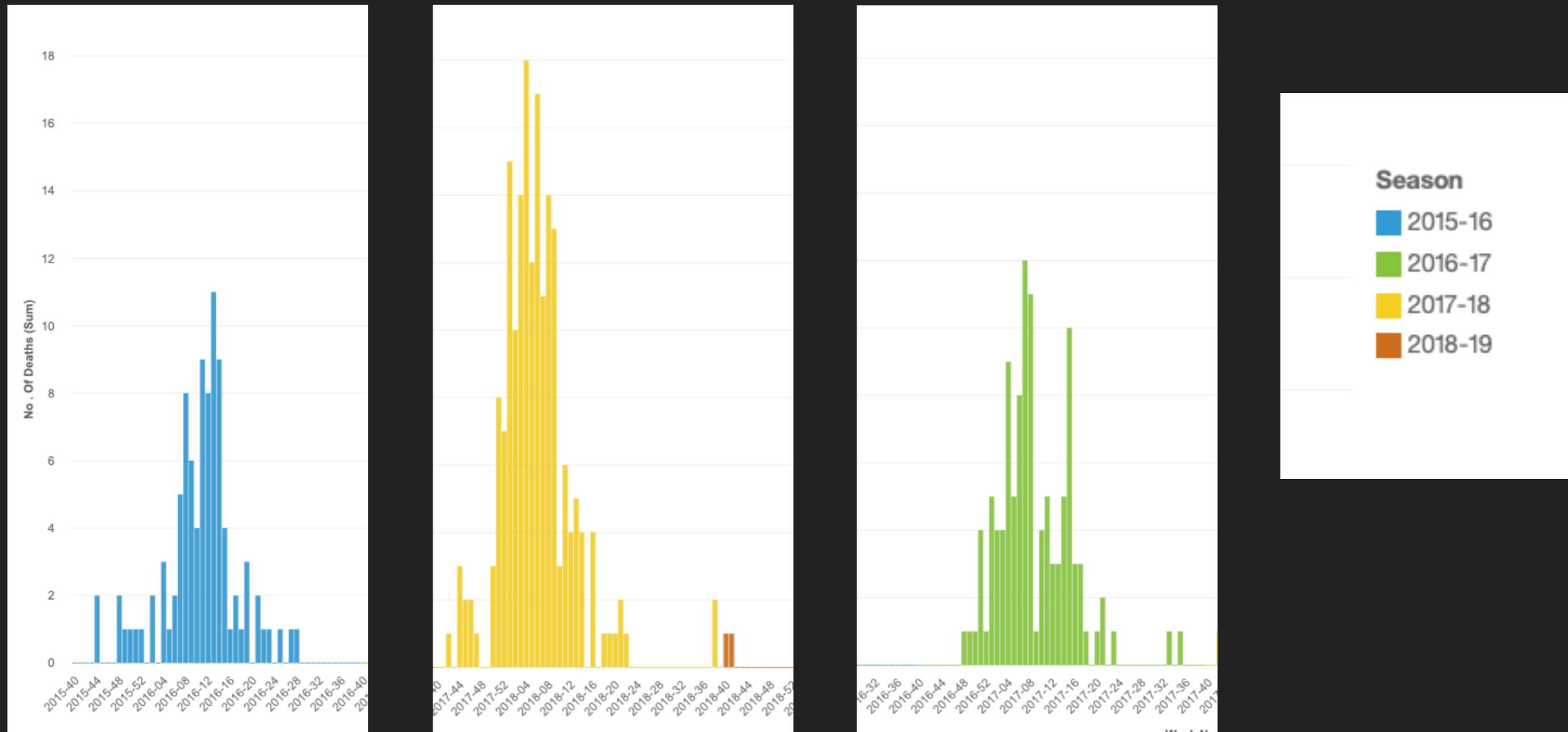
Activities mapped where darker state represents higher level of ILI activity compared to lighter states. (Level 1 - Level 10)

# PAEDIATRIC MORTALITY

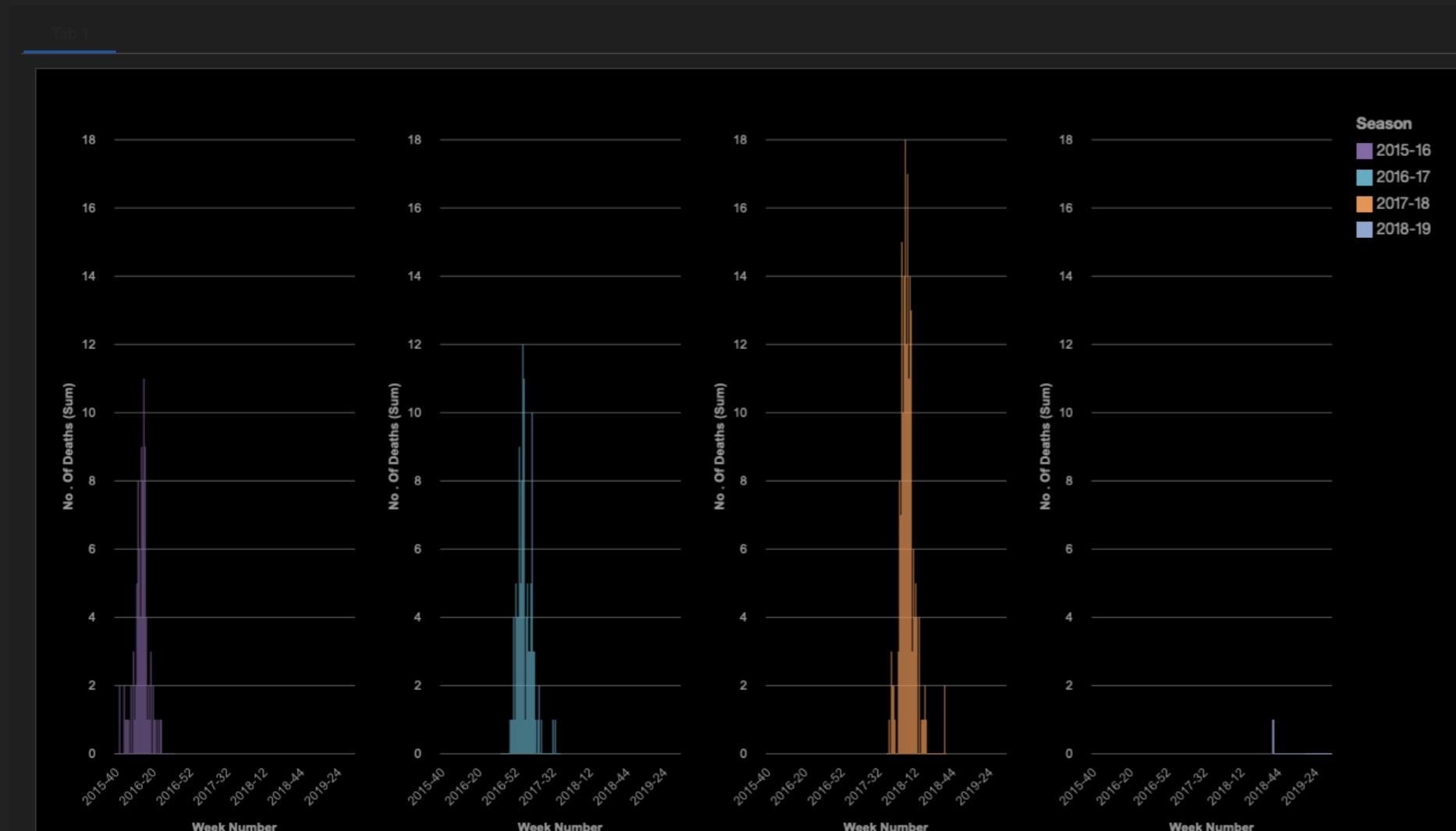
Child Mortality grouped by season 2015-2019 represented on Column Chart.



# EACH COLOR REPRESENTS DIFFERENT SEASON



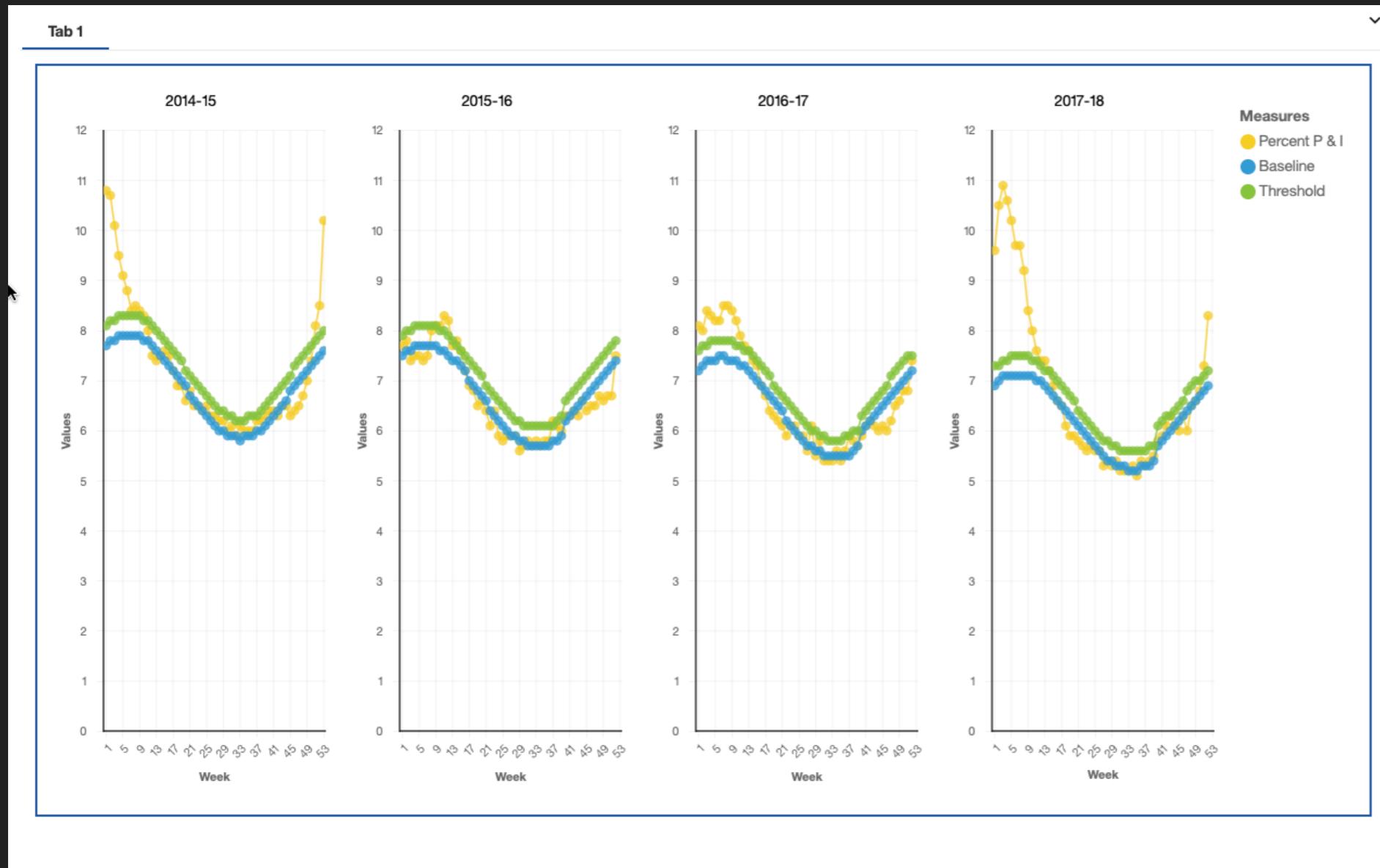
# SEASON-WISE VISUALIZATION



We can separate the visualization by seasons where it shows different charts for each season.

# P & I MORTALITY

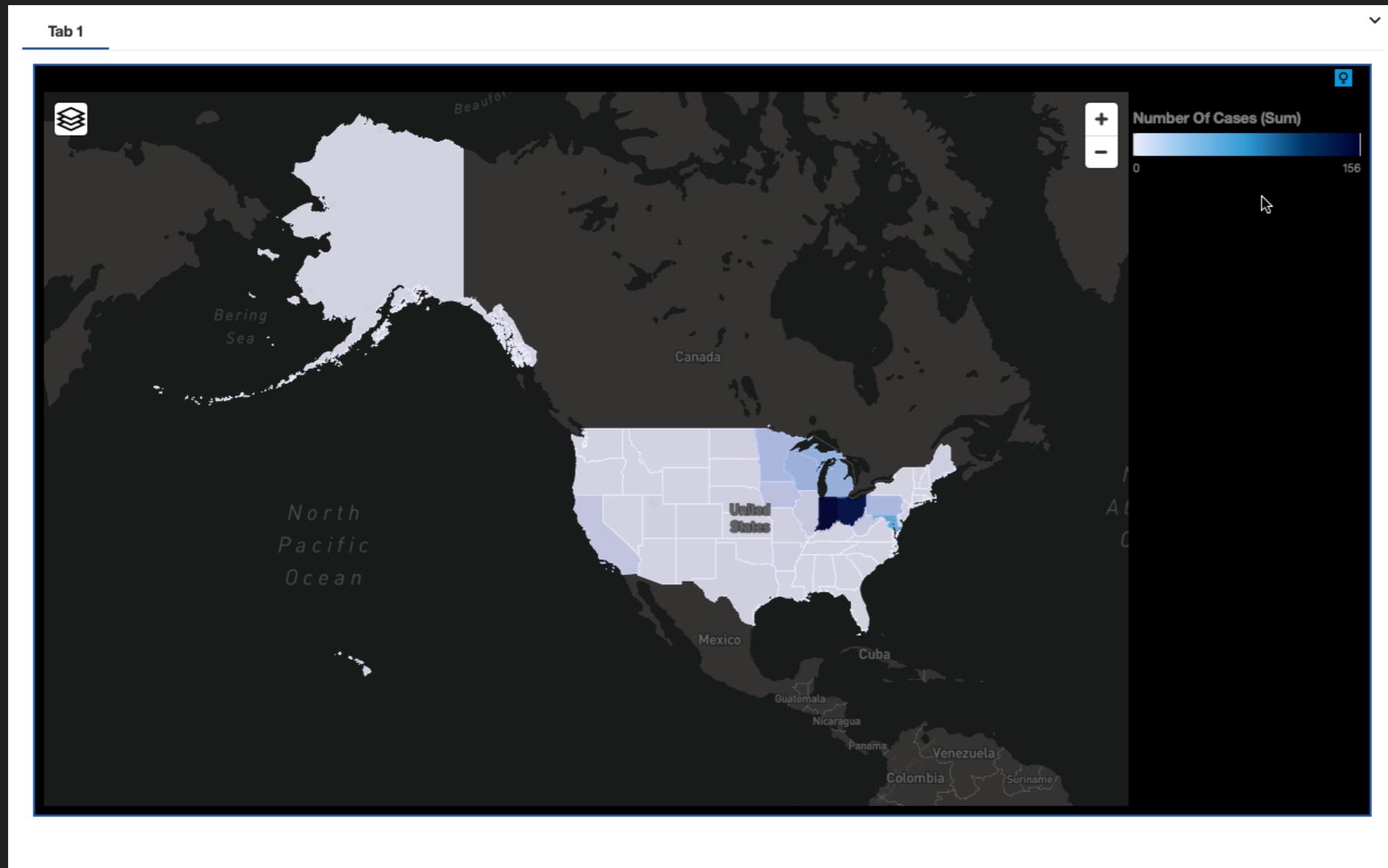
Surveillance of Death caused due to Pneumonia and Influenza represented on a Line Chart



Here each chart represents death caused by Pneumonia and Influenza from year 2014 - 2018.

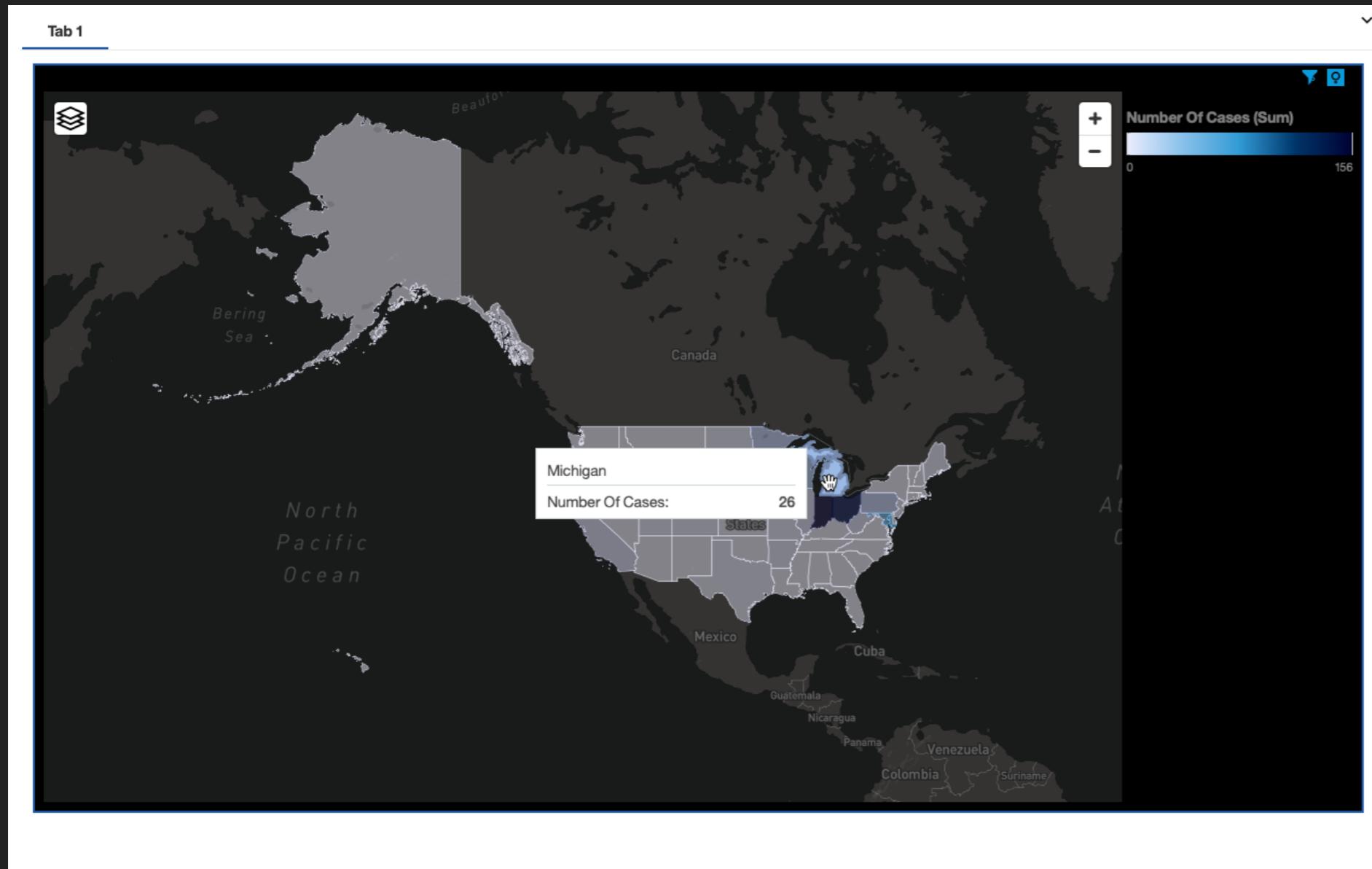
# NOVEL FLU A

## Novel Influenza A Virus Infections Cases by State



Here lighter region represents less number of cases while darker region with more number of cases.

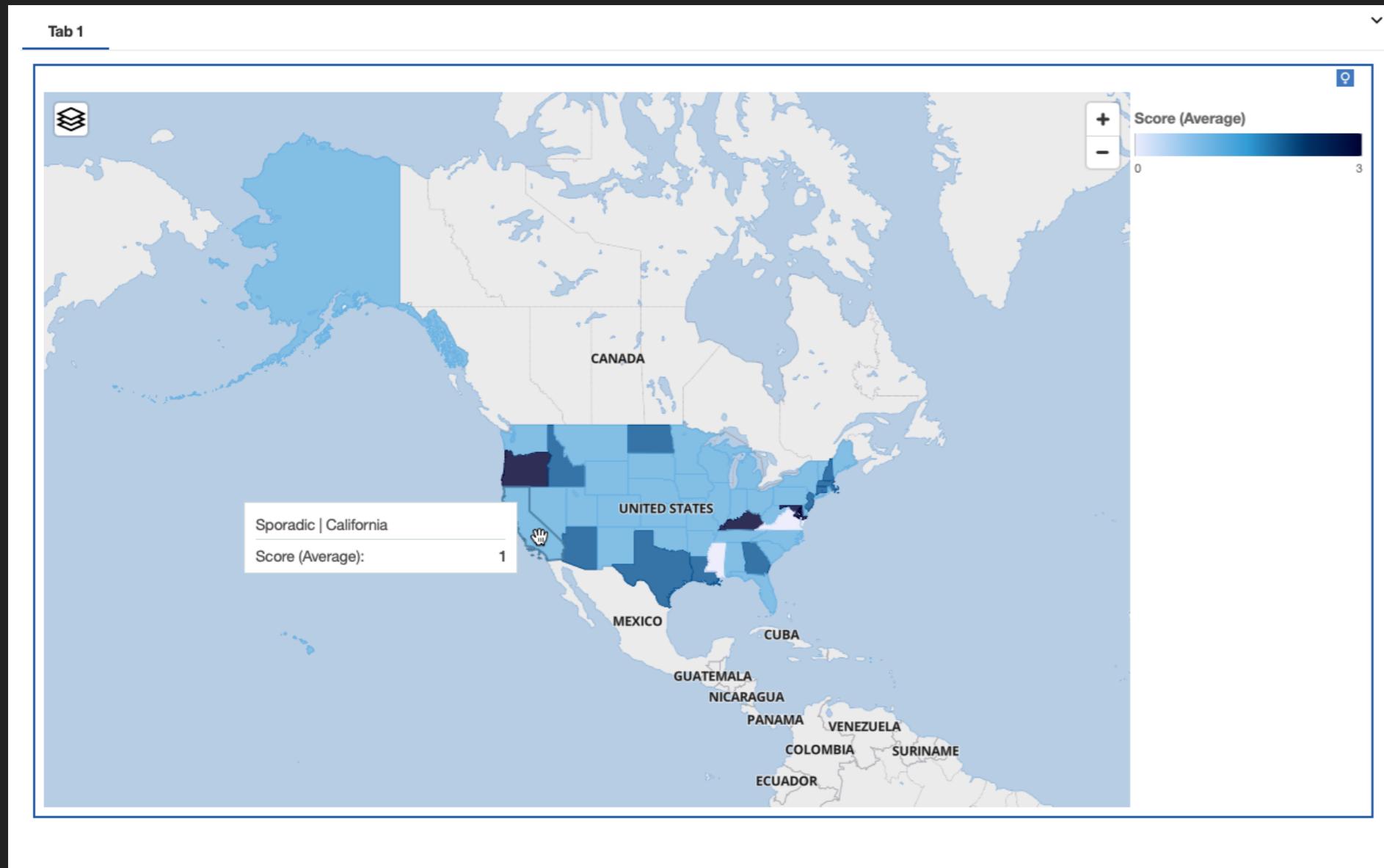
# HOVERING CURSOR REVEALS DETAILS



Hovering cursor on a specific region reveals more details like name of state and number of cases.

# GEOGRAPHIC SPREAD

State and Territorial Expert's Report on Geographical Spread of Influenza.



Here Light Blue Color Represents Sporadic Activity, Medium Dark Blue Represents Local Activity and Dark Blue Represents Regional Activity.

# CAPSTONE PROJECT ON DEBT RECOVERY ANALYSIS AND PREDICTION FOR FUTURE LENDING AND RECOVERY

---

PREPARED BY:  
MONAL GOSAI

# INTRODUCTION

- ▶ Analysis for Debt Recovery.
- ▶ Prediction for Debt Recovery.
- ▶ Prescription for Future Lending.



## DATA SET USED

- ▶ Historical Financial data of United States of America.
- ▶ Data set time line 2007 - 2017.
- ▶ Data set with 46k Rows and 150 Columns.
- ▶ Source: <https://www.kaggle.com/>

## TECHNOLOGIES USED

- ▶ Microsoft Azure Notebooks.
- ▶ Python 3.6 on Jupyter Notebooks.
  - ▶ Libraries: NumPy, Pandas, Series, MatPlotLib, Dataframe.
- ▶ Microsoft Power BI for visualizations.

## ANALYSIS AND CLEANING

- ▶ Numpy, Pandas, Dataframe Python libraries used for analysis.
- ▶ Columns with unnecessary values values were removed.
- ▶ Rows with 'NaN' / Null values were replaced with understandable numeric values.

```
In [4]: df.shape  
Out[4]: (42538, 151)
```



```
In [14]: df = df.dropna(how='all', axis=1)  
In [16]: df.shape  
Out[16]: (42538, 69)
```

## ANALYSIS AND CLEANING

- ▶ Character values were difficult to recognize by machine so they were converted to numeric.
- ▶ Example: column- loan\_status:

FullyPaid = 1,

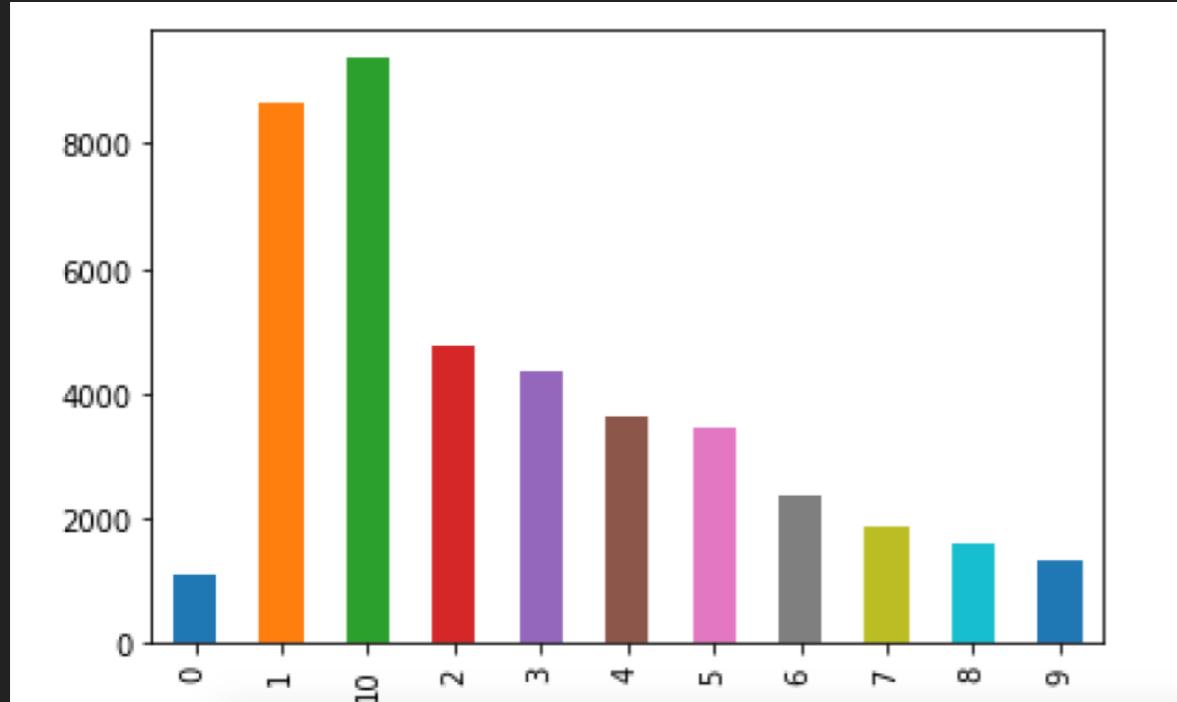
ChargedOff = 0.

- ▶ Same thing was performed for multiple columns.

## MEAN, MEDIAN

- ▶ Mean and Median played important role in analyzing the dataset.
- ▶ Classification or groups of costumes were formed using these functions.
- ▶ Example: column annualIncome:
- ▶  $0 - 15k = 0$ ,  $15k - 30k = 1$ ,  $30k - 50k = 2$ ,  $50k - 75k = 3$

# VISUALIZATIONS



A graph showing the relation between years of experience and income ratio.

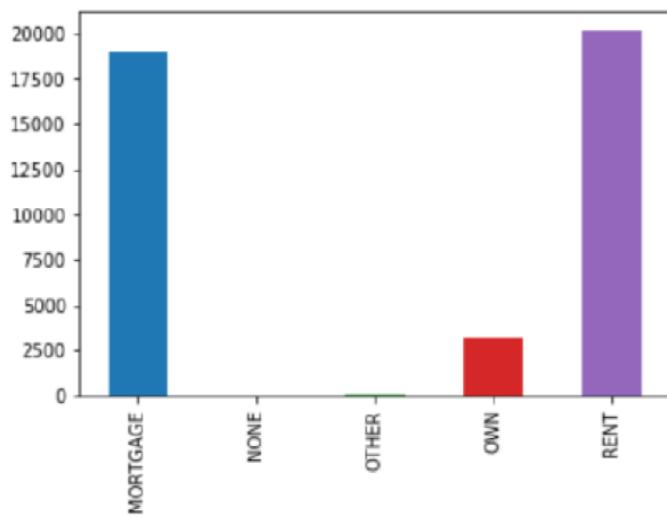


A graph showing the relation between Loan status according to level of income.

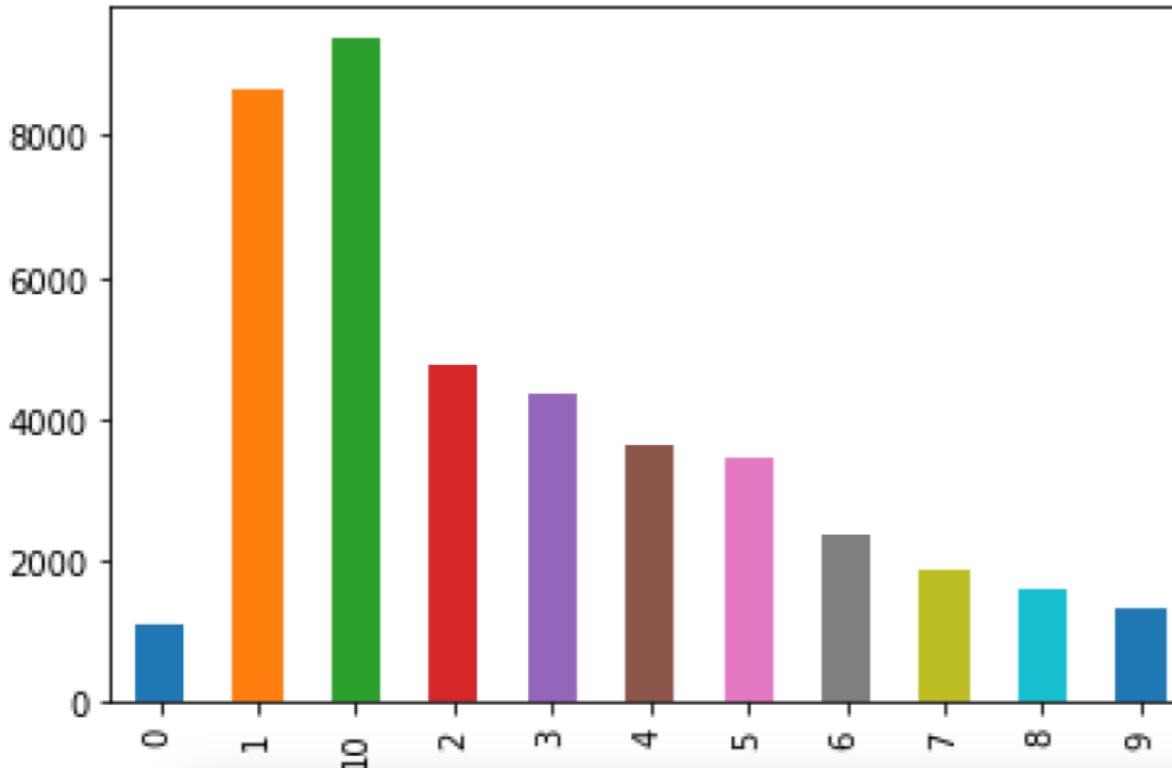
# VISUALIZATIONS

```
In [18]: df['home_ownership'].value_counts().sort_index().plot.bar()
```

```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x7f368c2e9b70>
```



A graph representing the home ownership related to individuals incomes



A graph representing the years of experience of employees and income status.

## PREDICTION AND MODELLING

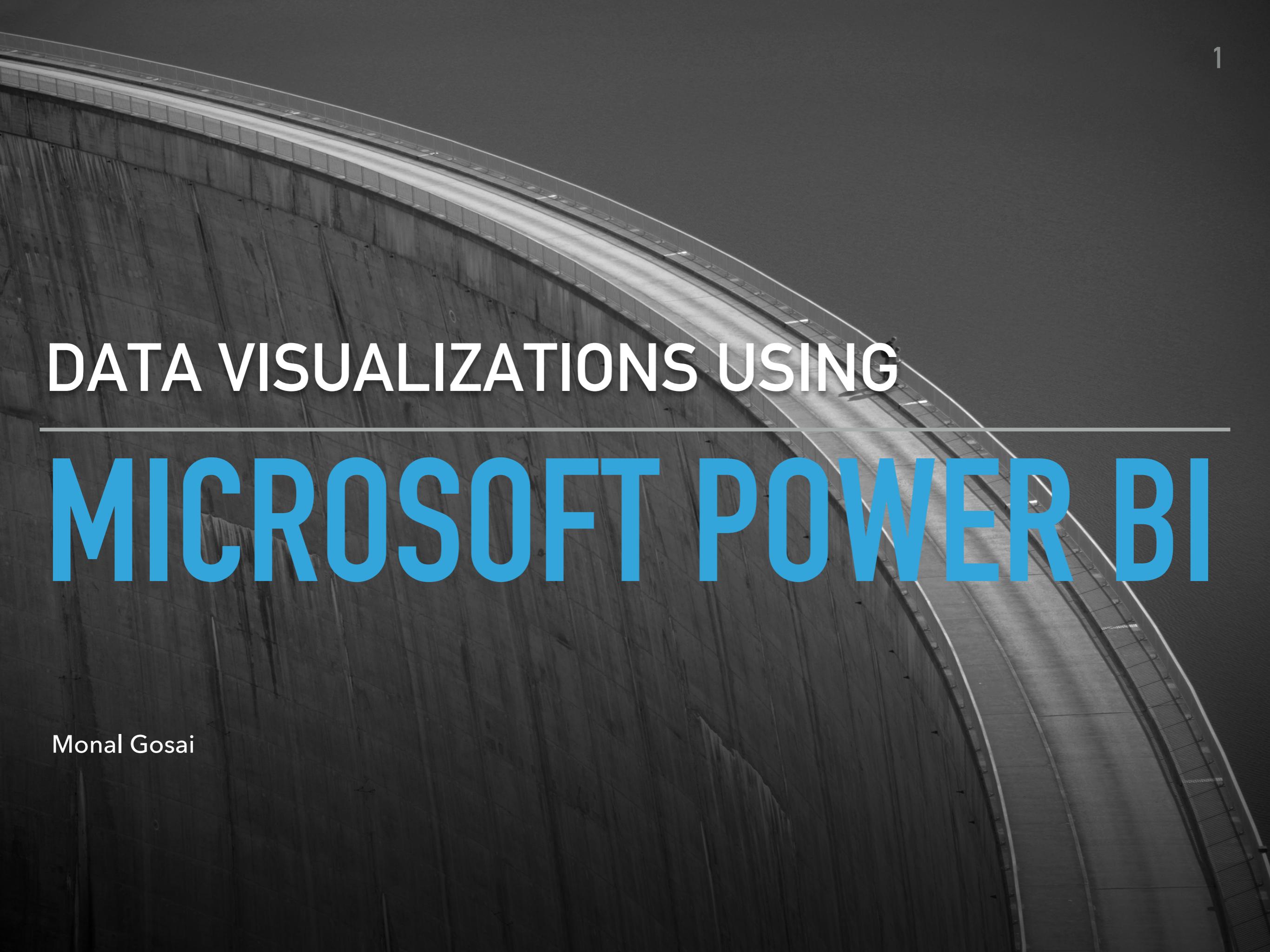
- ▶ Sklearn - KFold, SVM, Logistic Regression, RandomForestClassifier, Naive Byes Classifier, XGBoost.
- ▶ Accuracy:
- ▶ LogisticRegression: 0.998
- ▶ RandomForest: 0.997
- ▶ NaiveByes: 0.842
- ▶ XGBoost: 0.973

## PRESCRIPTION

- ▶ Logistic Regression: [1,0,1,1,1]
- ▶ Random Forest: [1,0,1,1,1]
- ▶ Naive Bayes: [1,0,1,1,1]
- ▶ XGBoost: [0,0,1,0,1]

## PRESCRIPTION

- ▶ According to all the models used on the dataset for this project. The probability of amount getting charged off was notably lower compared to the paid off amount.
- ▶ Still the results are mainly dependant on the income group of the clients.

The background of the slide features a high-angle, black and white photograph of a massive concrete dam. The dam's curved, ribbed structure dominates the frame, with a long, straight concrete walkway or roadway running along its top edge. The surrounding terrain appears rugged and arid.

# DATA VISUALIZATIONS USING MICROSOFT POWER BI

Monal Gosai

# SELECT VISUALIZATION TYPE AND PLOT COLUMNS

The screenshot shows the Power BI visualization pane with the 'Values' tab selected. The pane includes a search bar, a list of data fields from the 'USCS\_SurvivalAllC...' dataset, and sections for page and report level filters, as well as drillthrough settings.

**Visualizations**

Show/hide pane >

Search

USCS\_SurvivalAllC...

- Σ 5-yearRelativ...
- CancerType
- Σ Ici
- Race
- Sex
- Σ uci

VALUES

Add data fields here

FILTERS

Page level filters

Drag data fields here

Report level filters

Drag data fields here

DRILLTHROUGH

Keep all filters

Off

Drag drillthrough fields here

Page Information  
Page Size  
Page Background  
Wallpaper

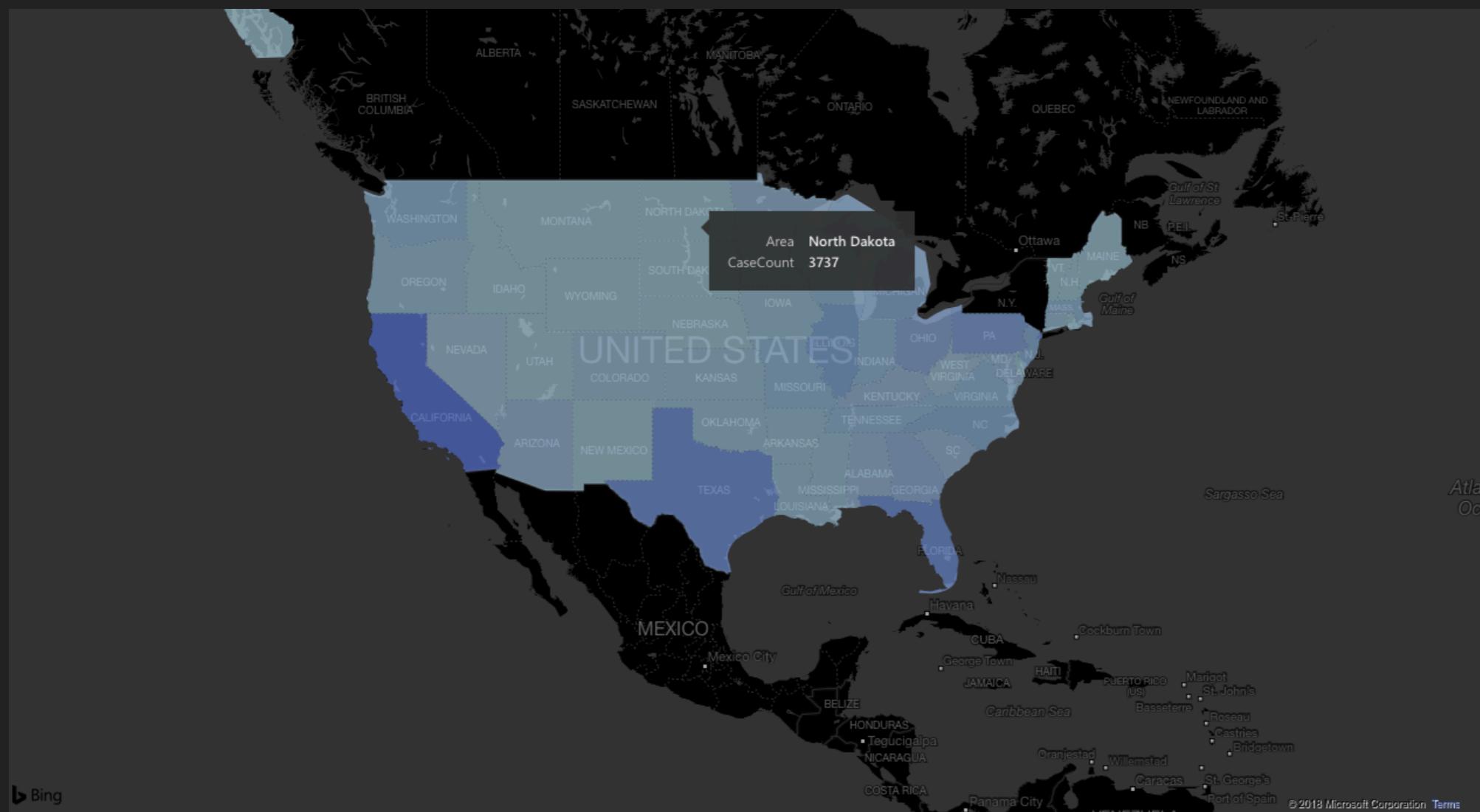
## DATA SET USED

### USA CANCER STATISTICS

- ▶ CANCER CASES AND DEATHS OVERVIEW
- ▶ RATE OF NEW CANCERS
- ▶ CANCER BURDEN: ALABAMA
- ▶ SURVIVAL: 5-YEAR RELATIVE SURVIVALNOVEL FLU A

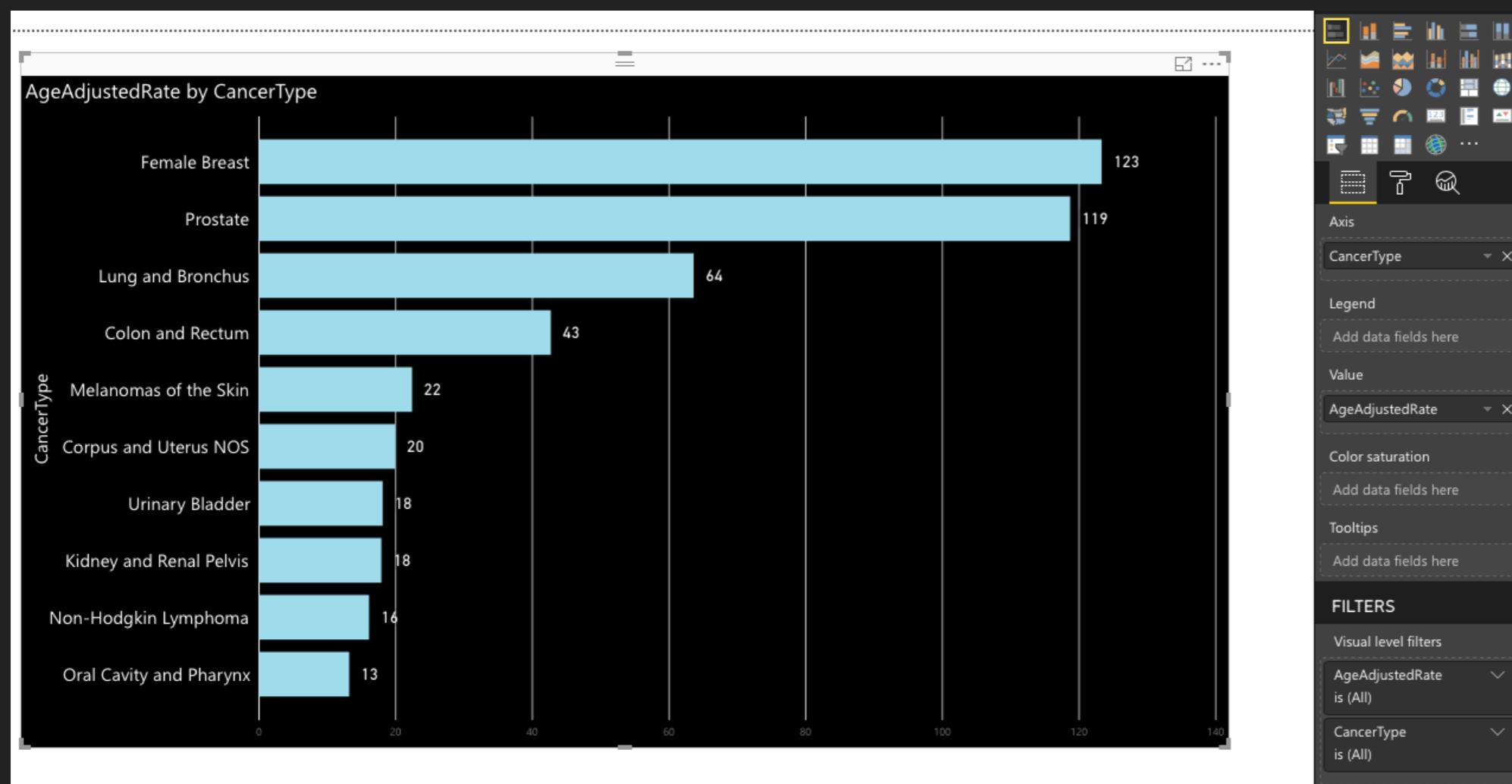
# CANCER CASES AND DEATHS OVERVIEW

All Types of Cancer, All Ages, All Races/Ethnicities, Male and Female  
Rate per 100,000 people



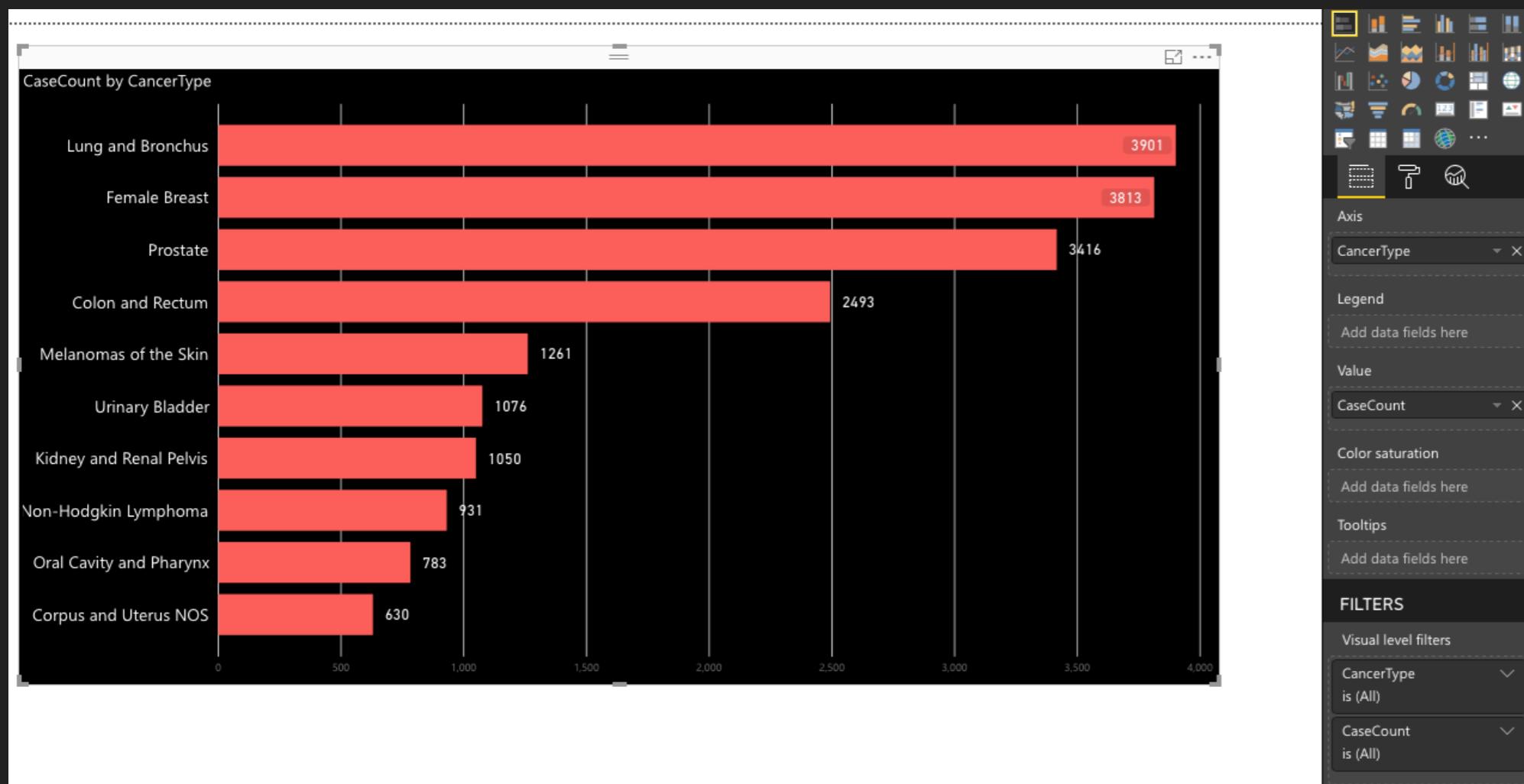
# CANCER CASES AND DEATHS OVERVIEW

## Rates of New Cancer Cases



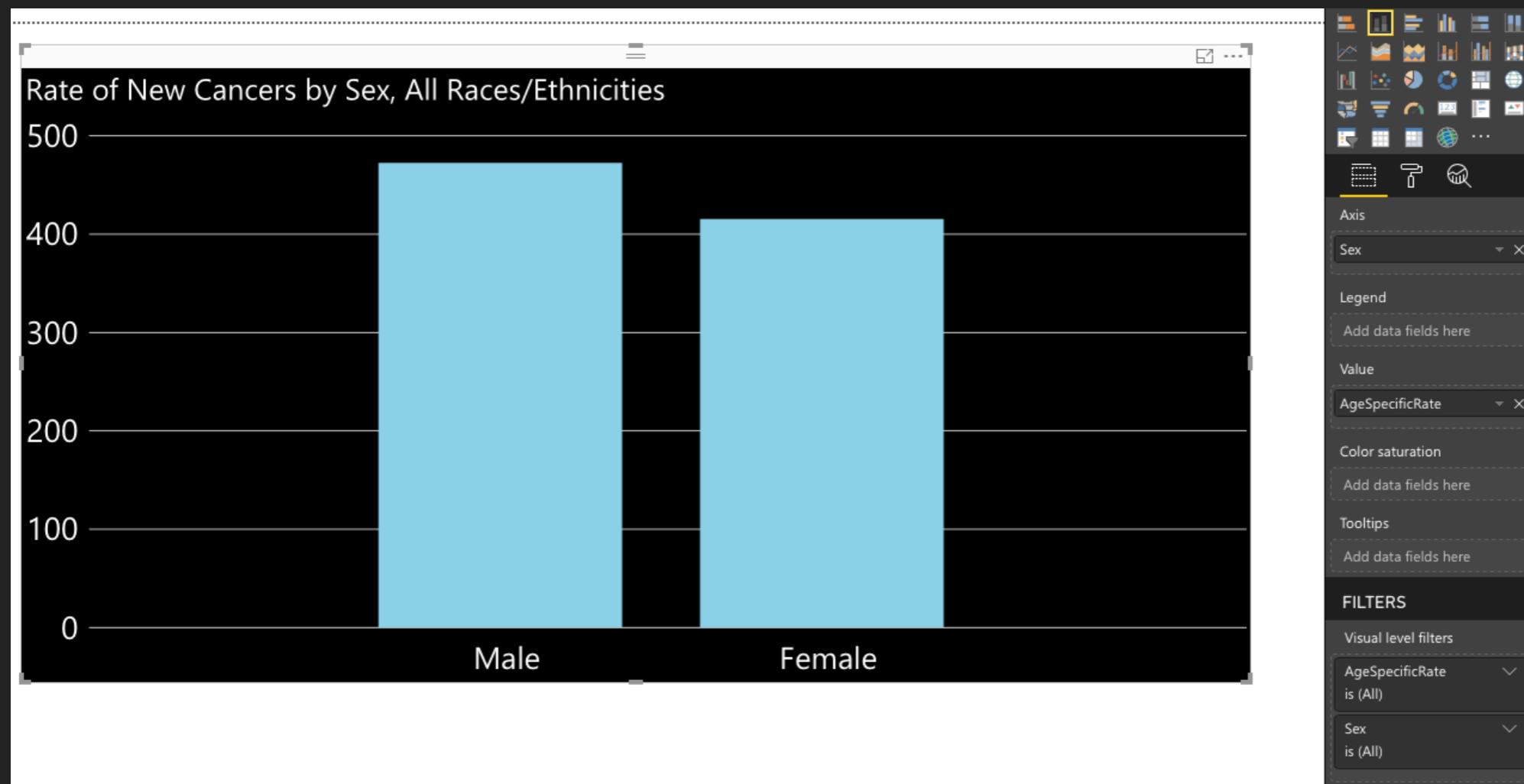
# CANCER CASES AND DEATHS OVERVIEW

## Rates of Cancer Deaths



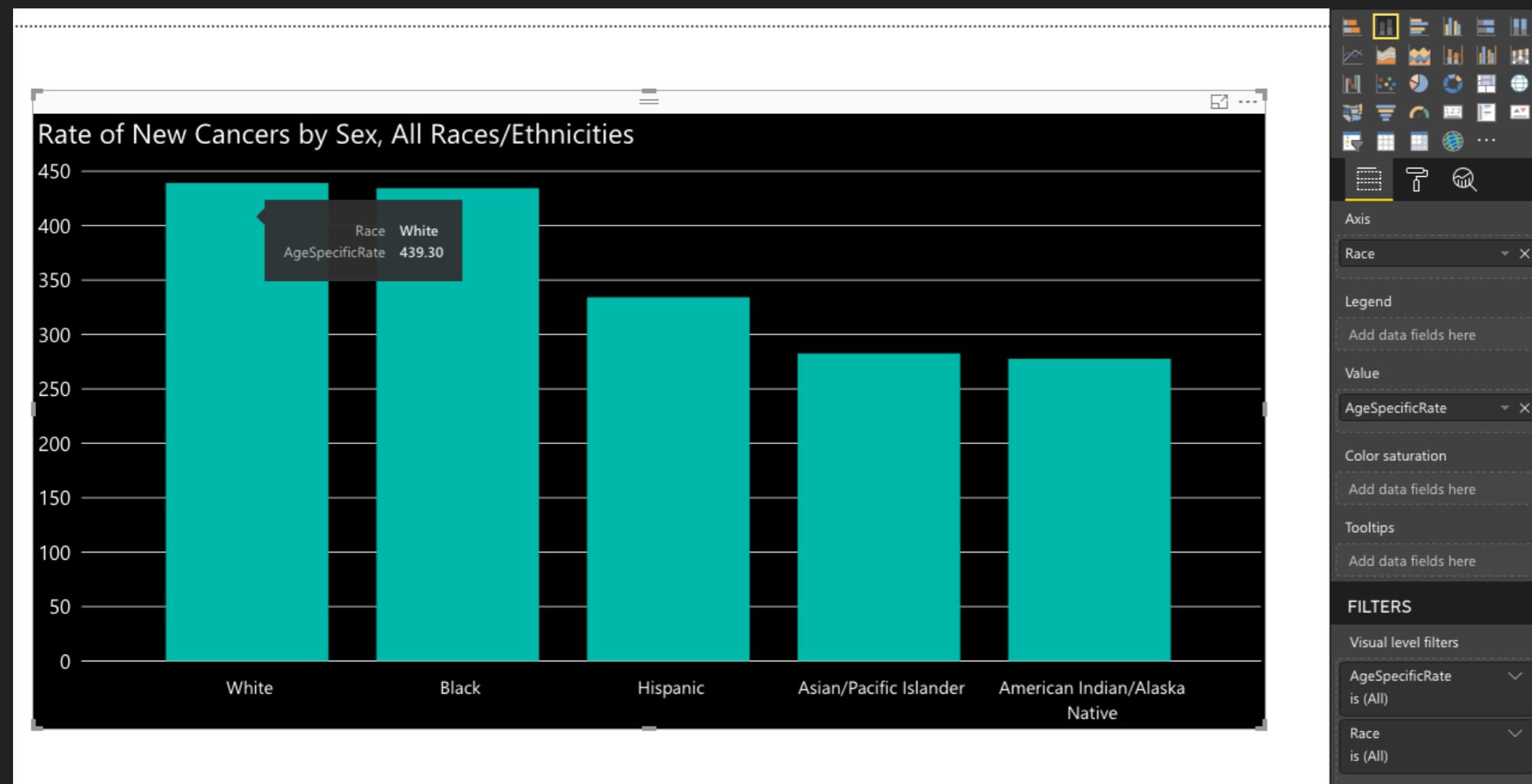
# RATE OF NEW CANCERS

Rate of new cancers by sex



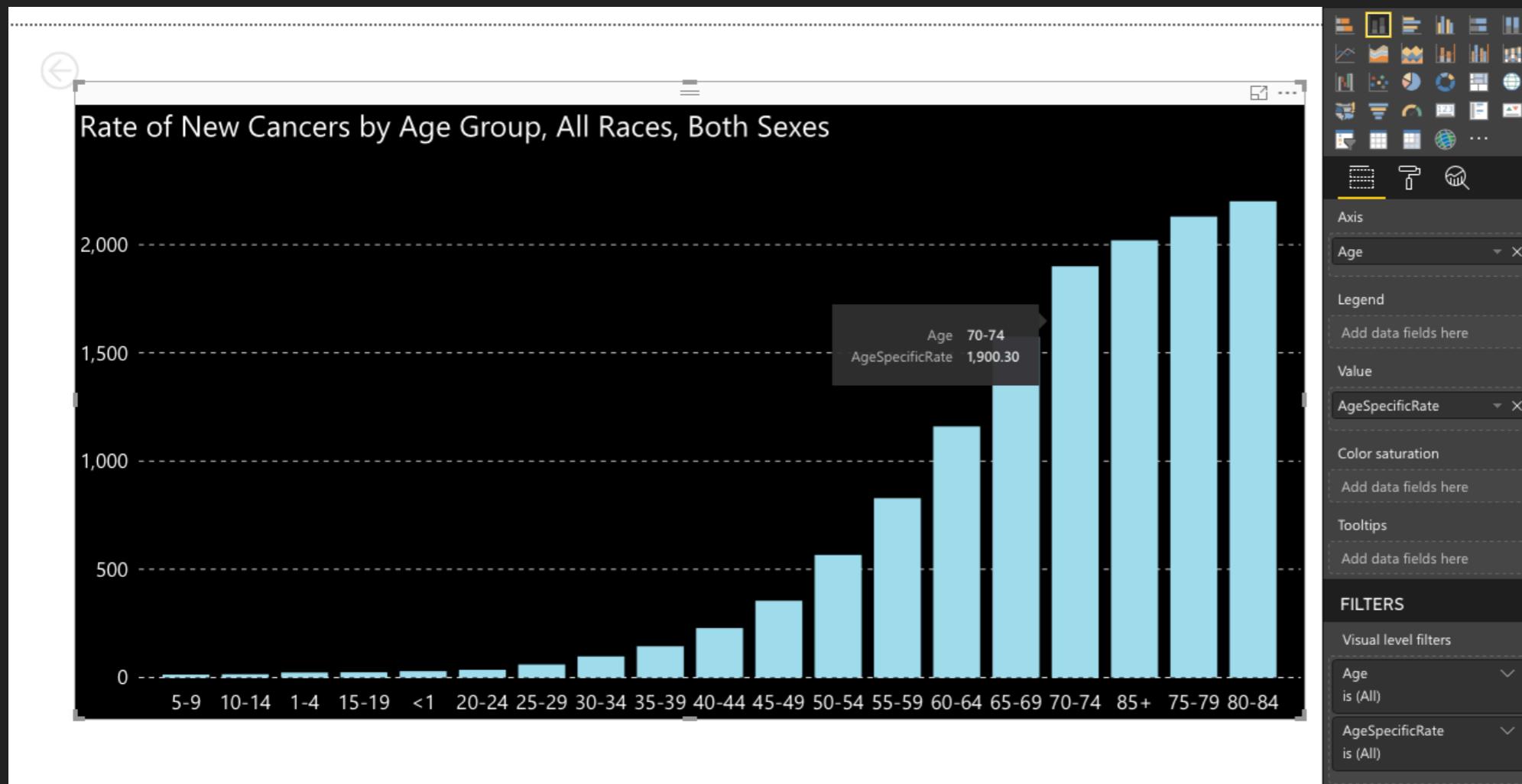
# RATE OF NEW CANCER

Rate of new cancers by Race



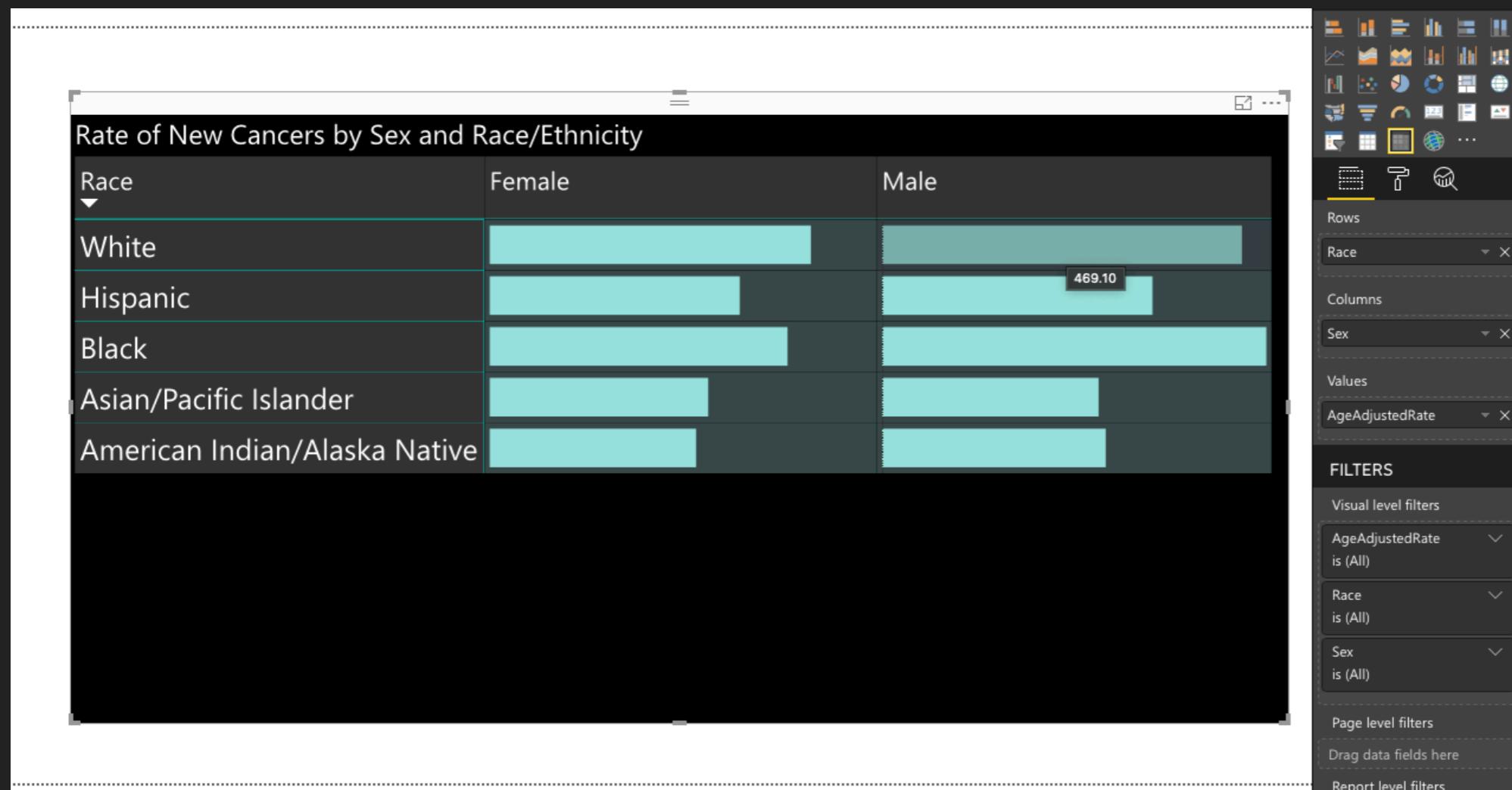
# RATE OF NEW CANCERS

## Rate of new cancers by Age Group



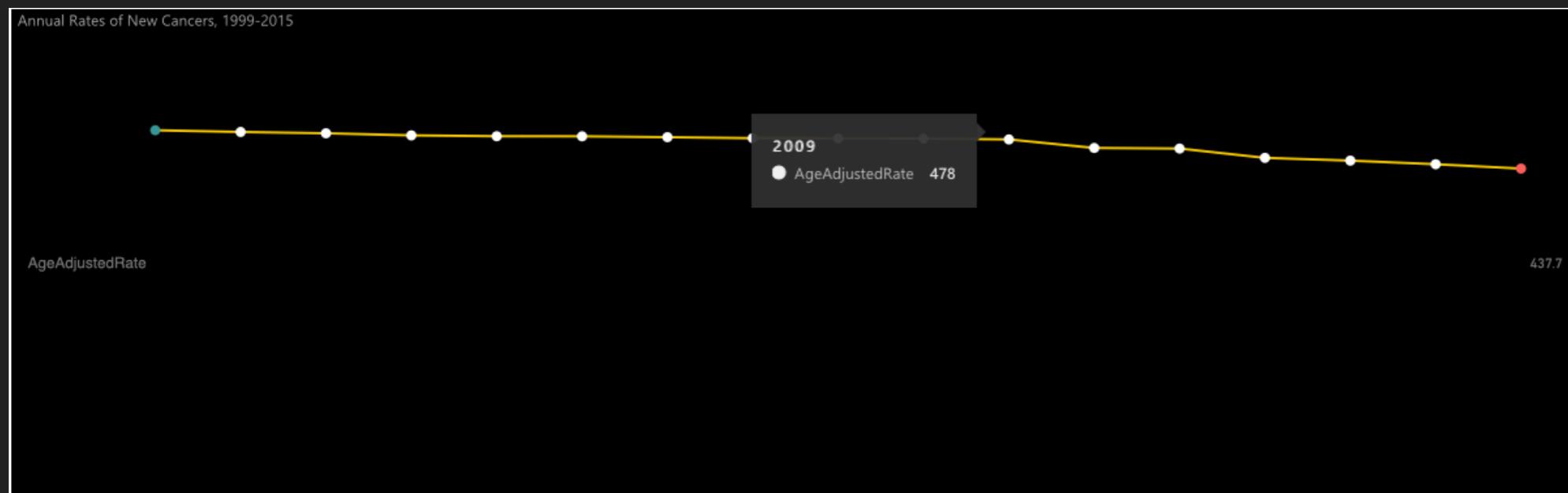
# RATE OF NEW CANCERS

Rate of new cancers by Sex and Race



# CHANGE OVER TIME / TRENDS

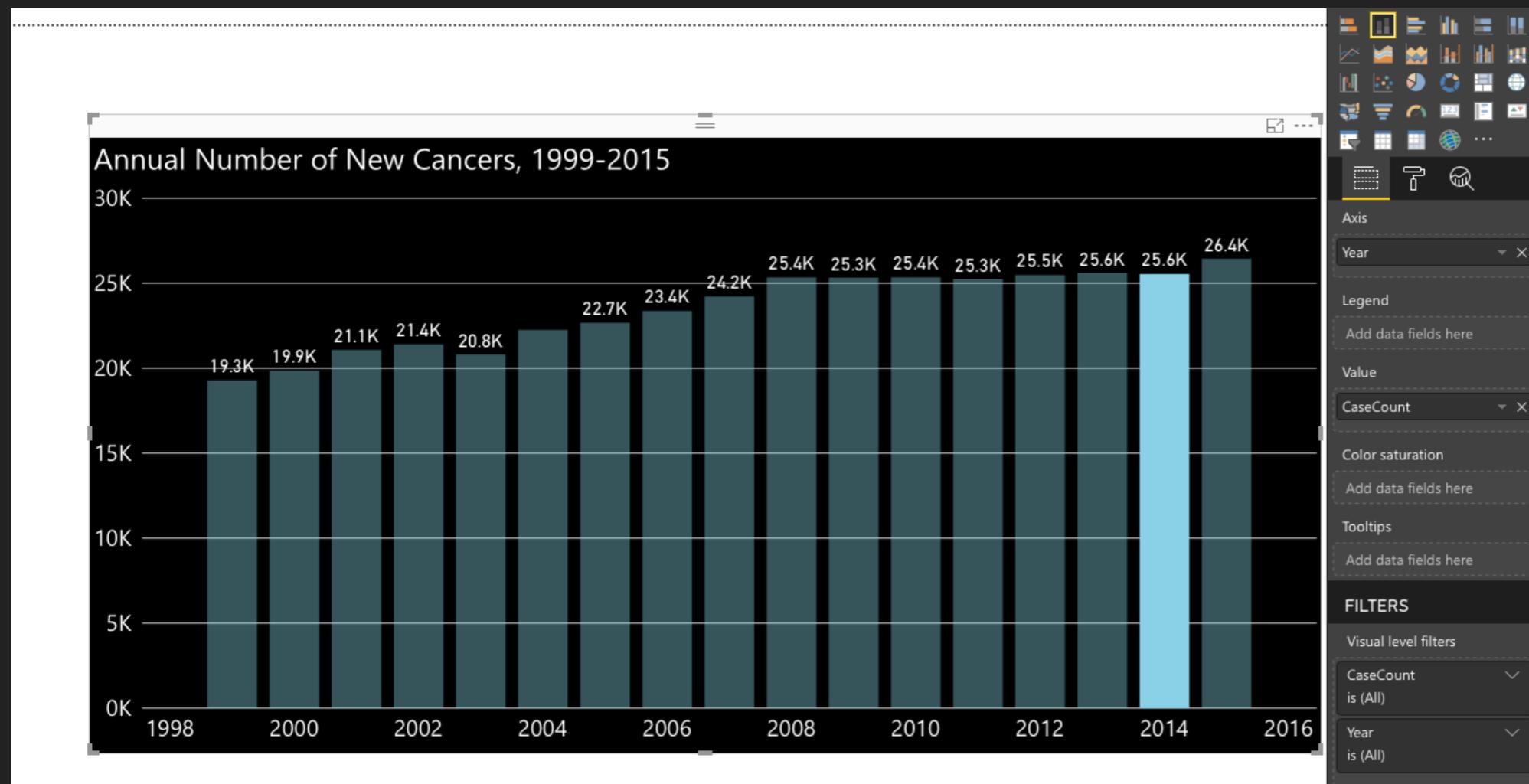
Annual Rate of New Cancers 1999 - 2015



For Dotted Line Chart we need to import custom library sparkline.

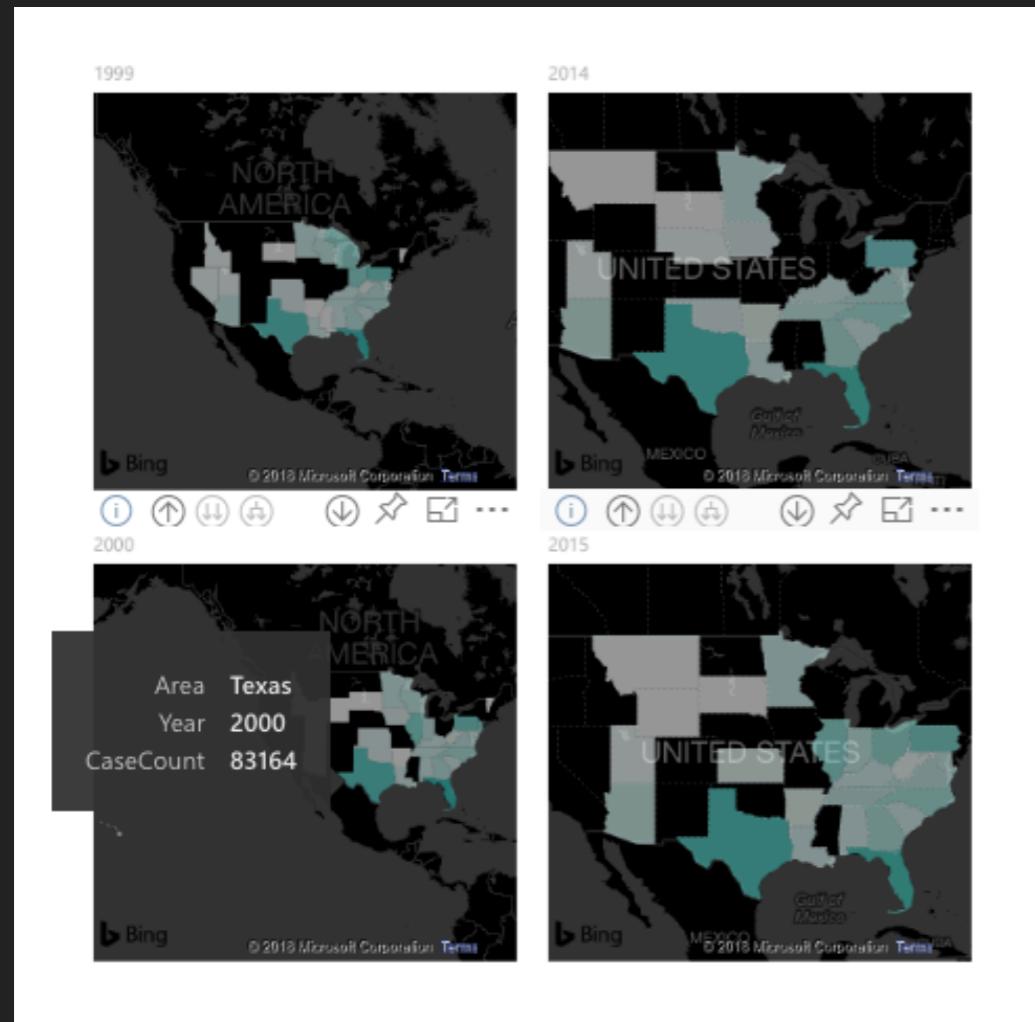
# CHANGE OVER TIME / TRENDS

Annual Numbers of New Cancers 1999 - 2015



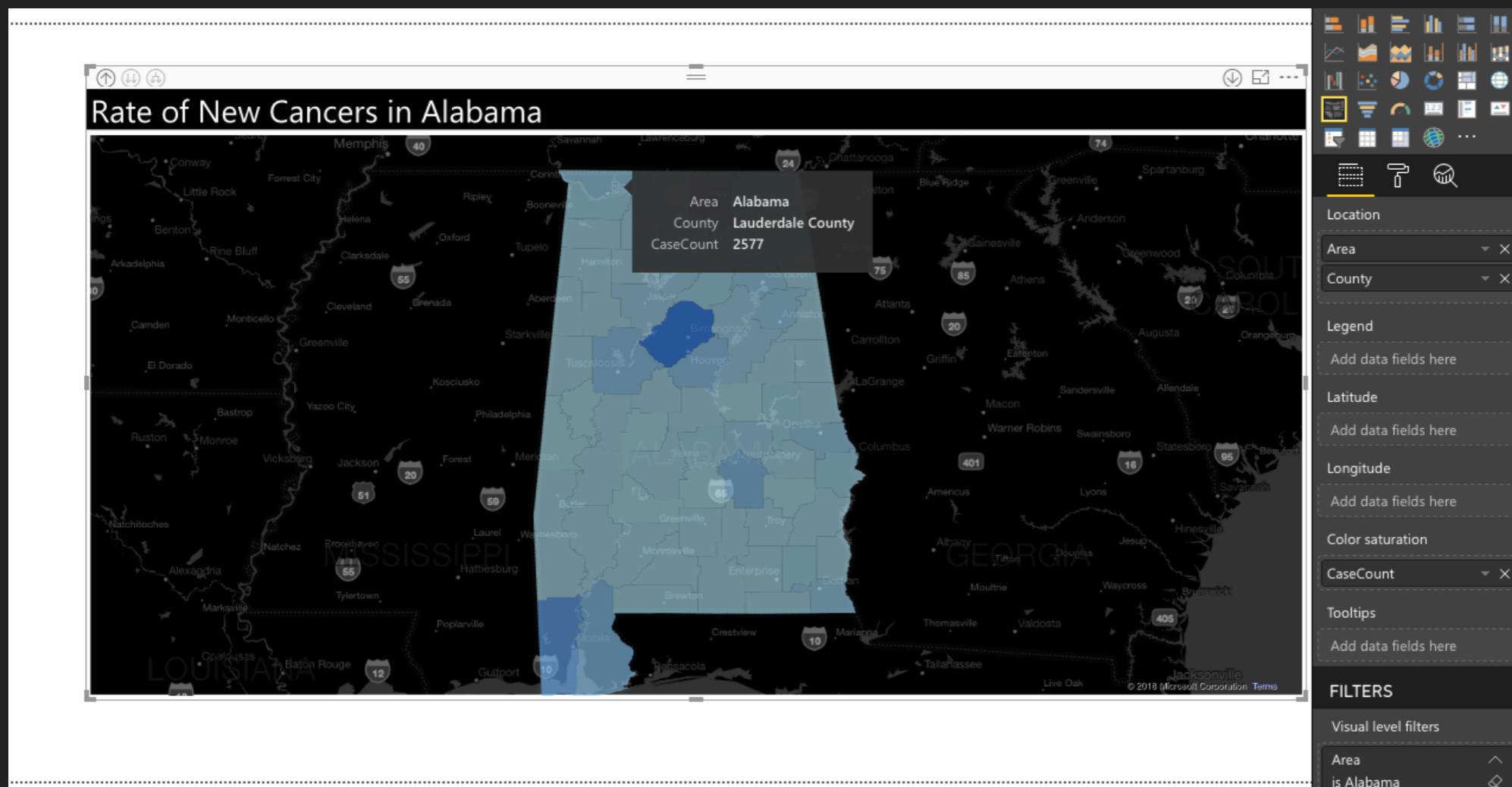
# CHANGE OVER TIME / TRENDS

Nation Wide Numbers of New Cancers 1999, 2000, 2014, 2015



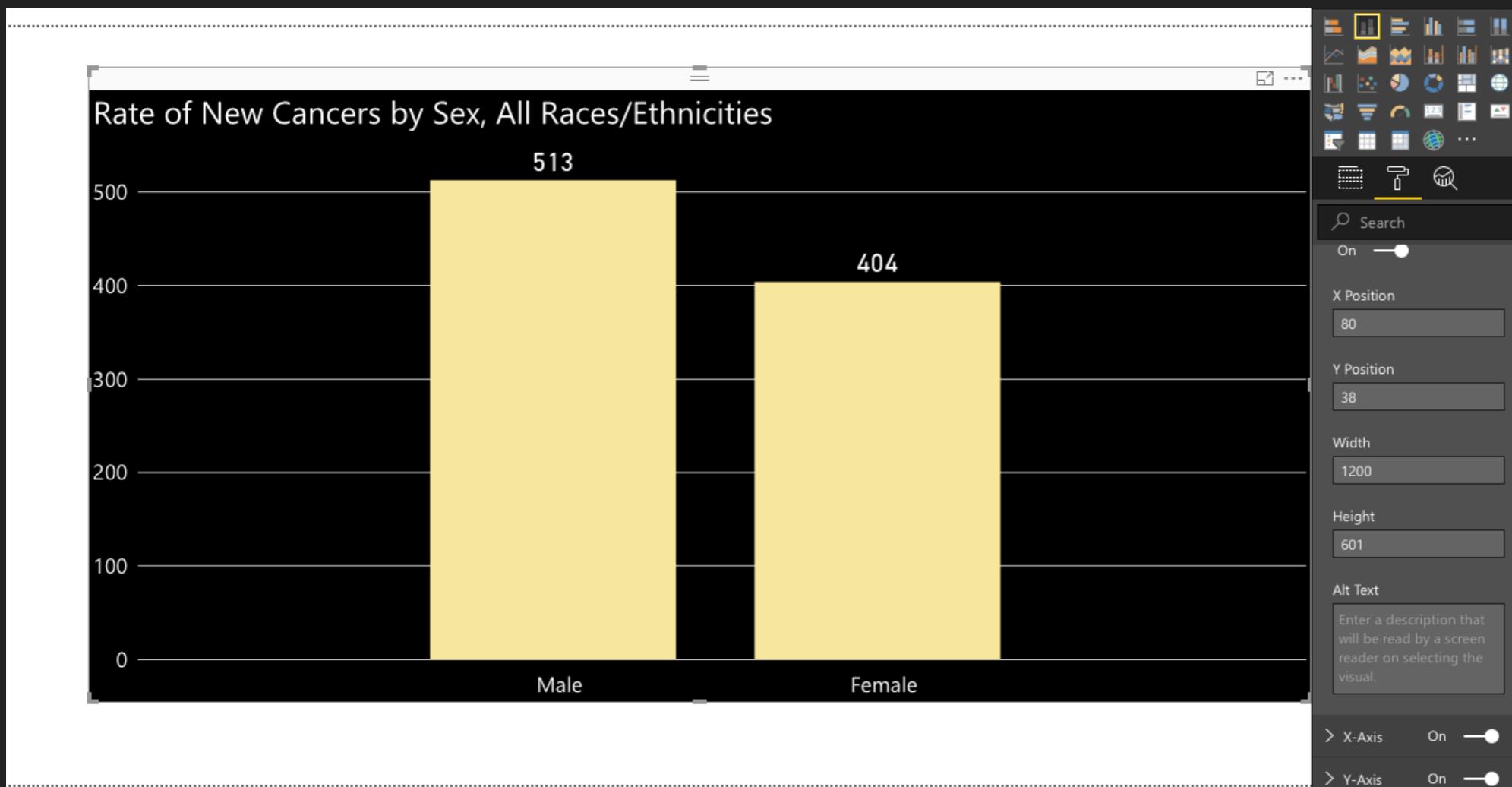
# STATE STATISTICS ALABAMA

## Rate of New Cancers in Alabama



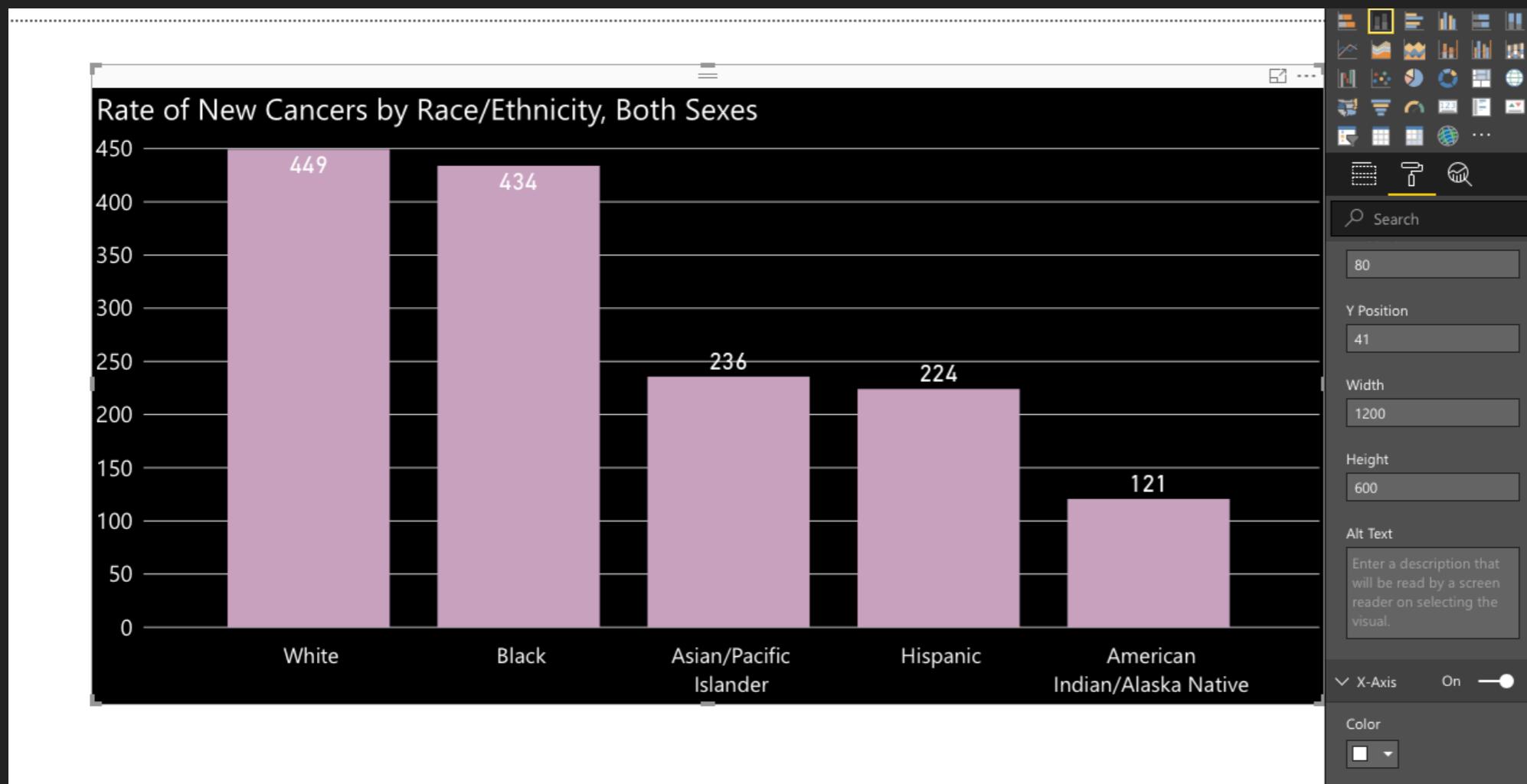
# STATE STATISTICS ALABAMA

## Rate of New Cancers by Sex



# STATE STATISTICS ALABAMA

## Rate of New Cancers by Race



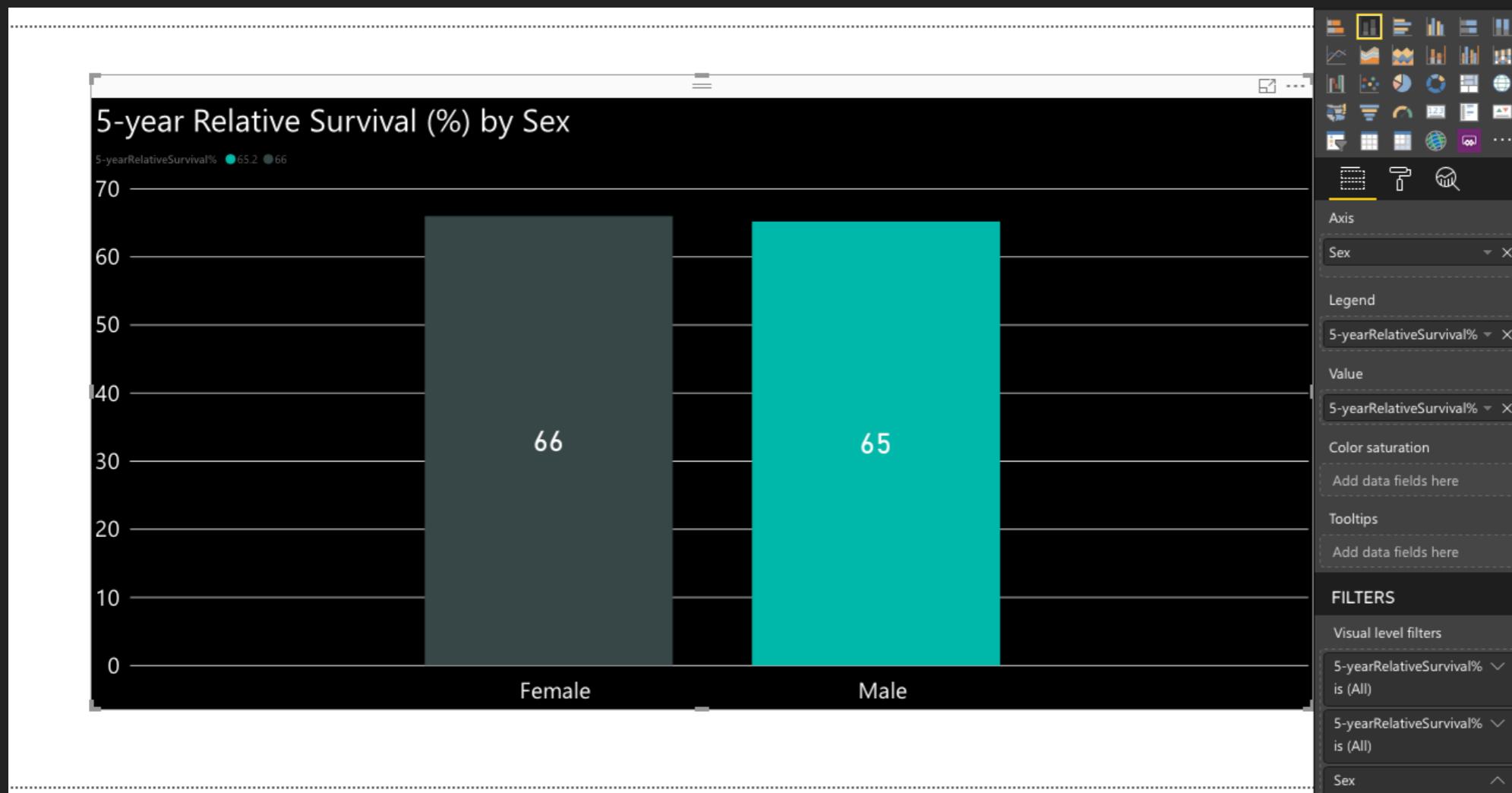
# STATE STATISTICS ALABAMA

## Rate of New Cancers by Race / Sex



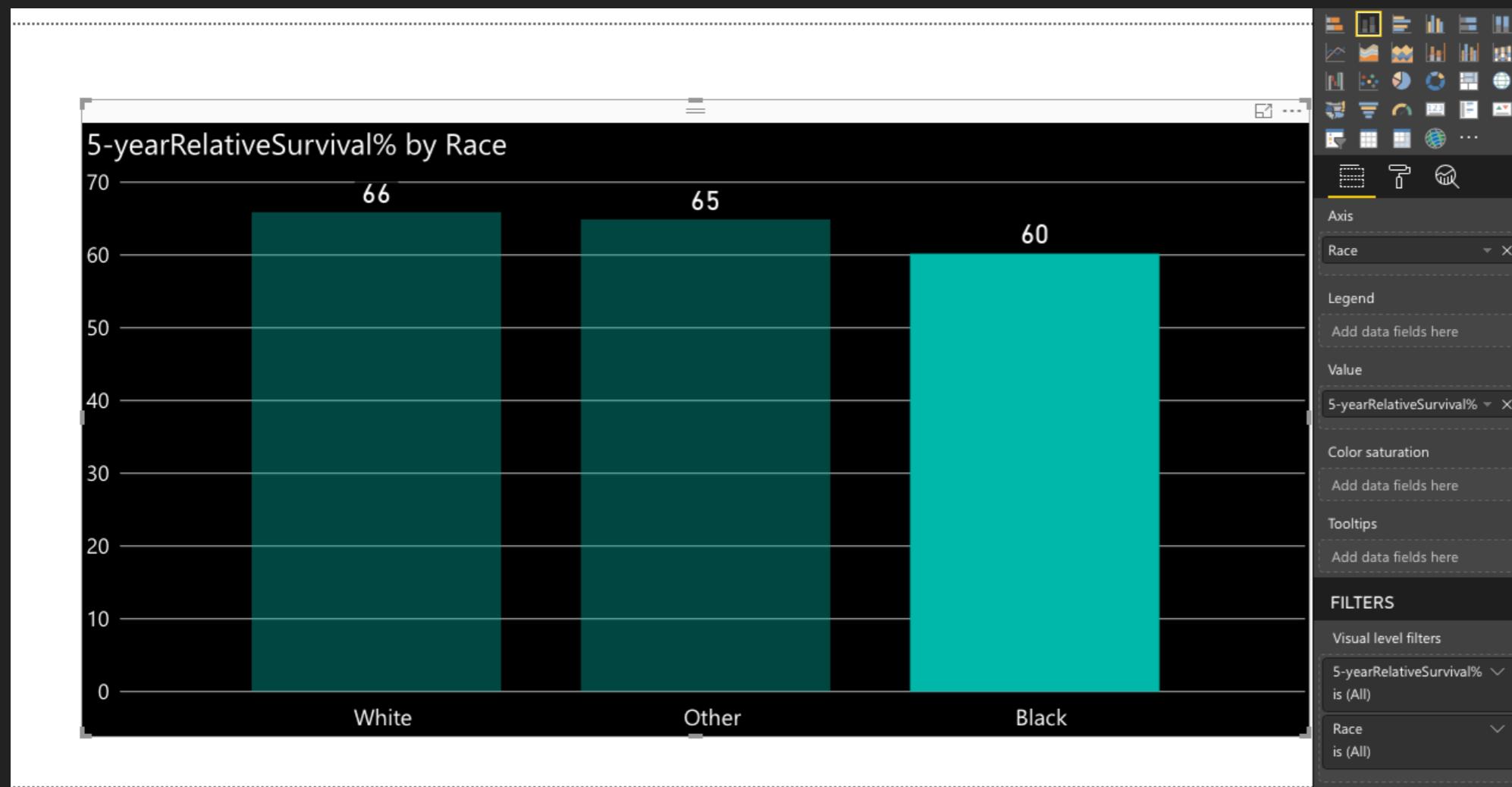
# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Sex



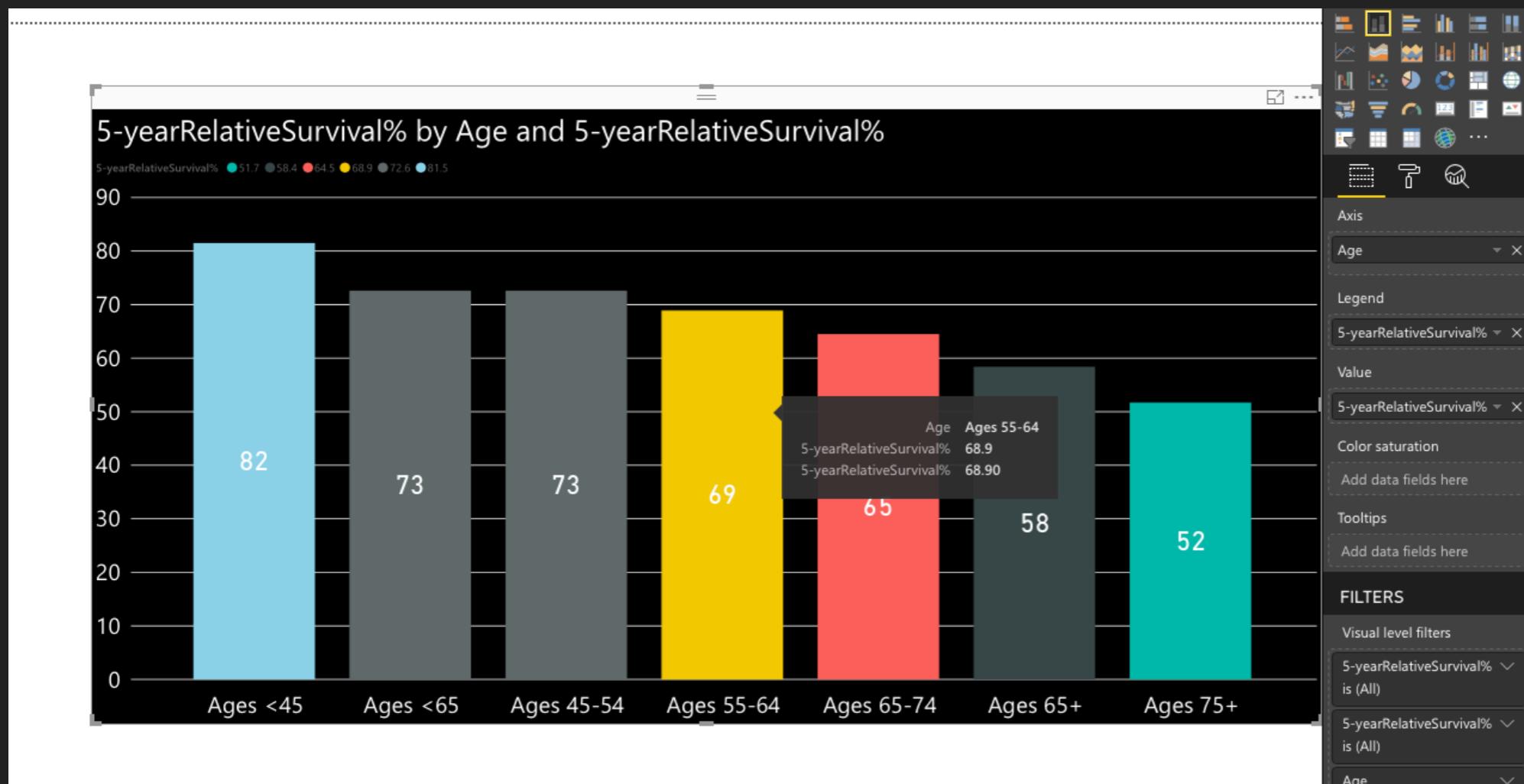
# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Race



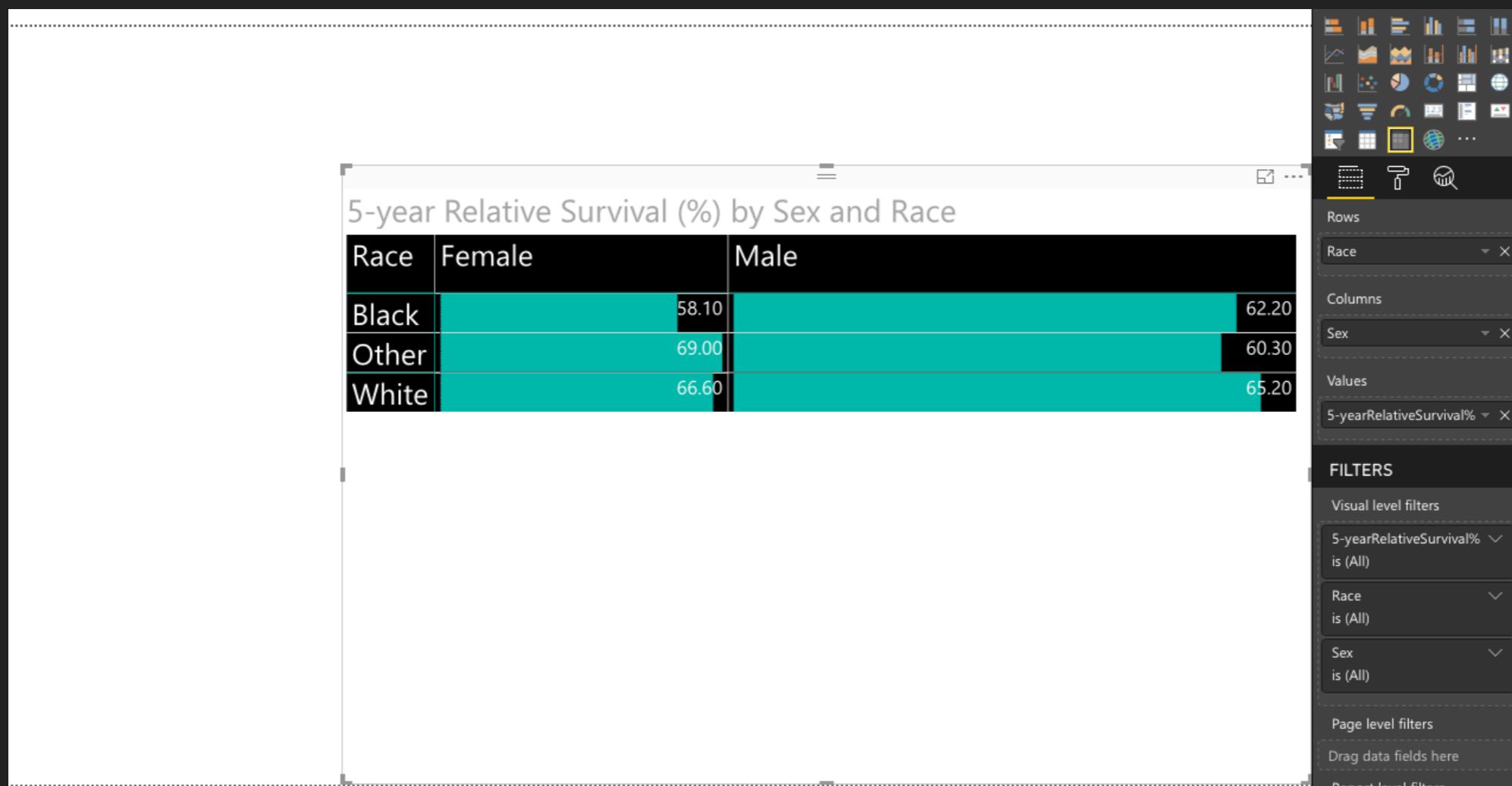
# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Age Group



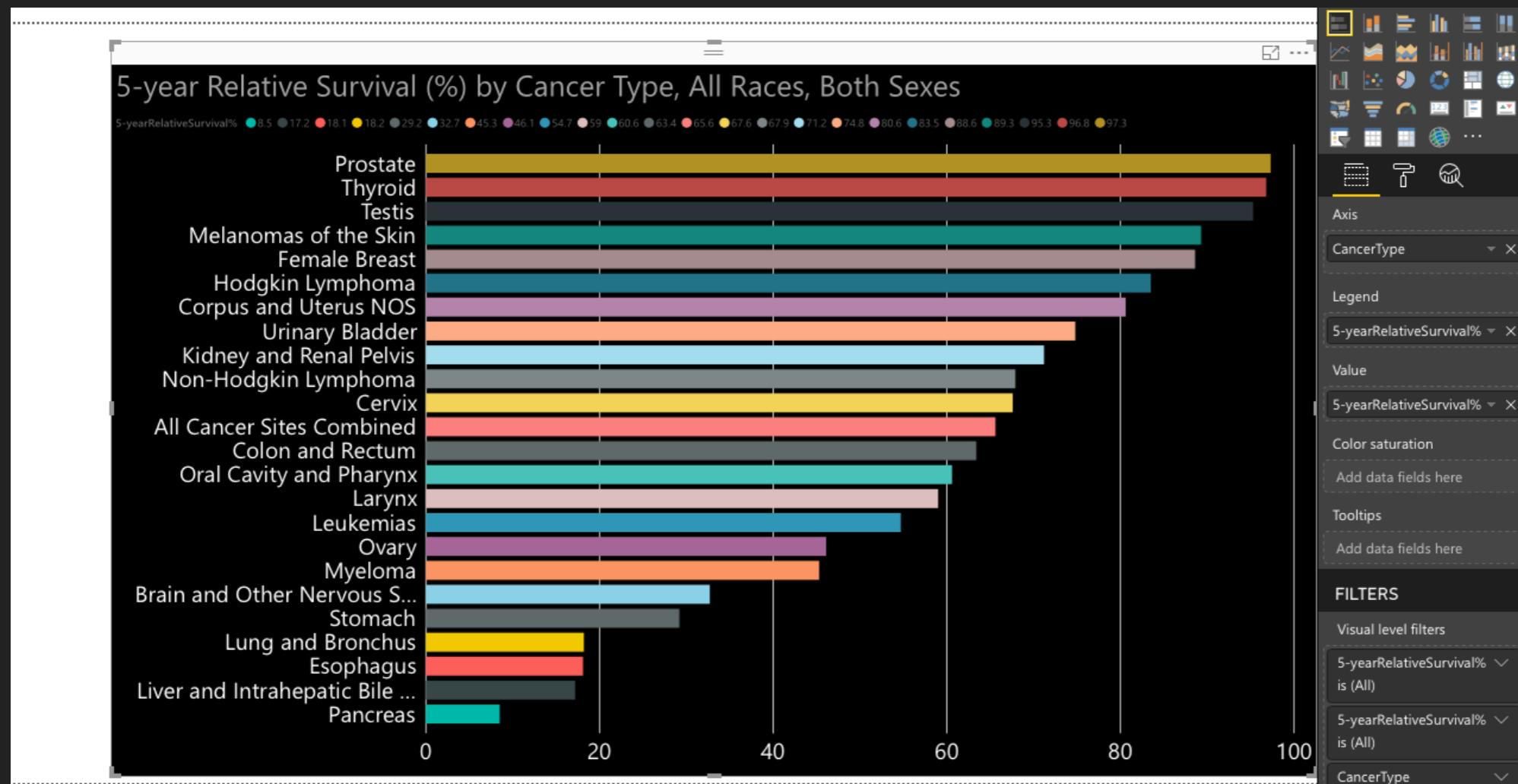
# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Sex / Race



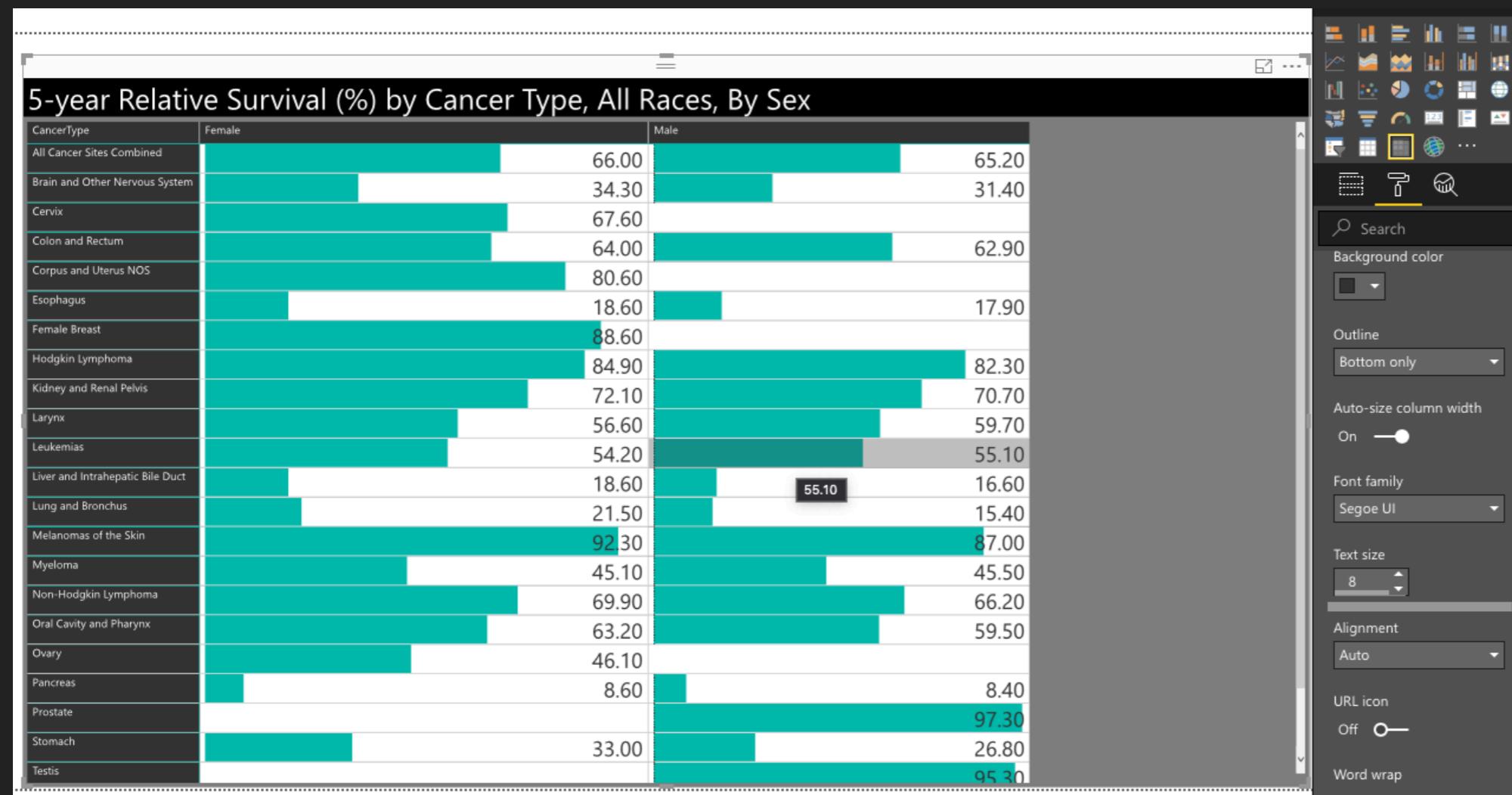
# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Cancer Type



# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Cancer Type / Sex



# SURVIVAL: 5 YEAR RELATIVE SURVIVAL

## 5 Year Relative Survival by Cancer Type / Sex / Race

