

# Exercise: Diabetes Analysis

Download and access all the available datasets for this exercise from the folder *datasets* in module 2 files.

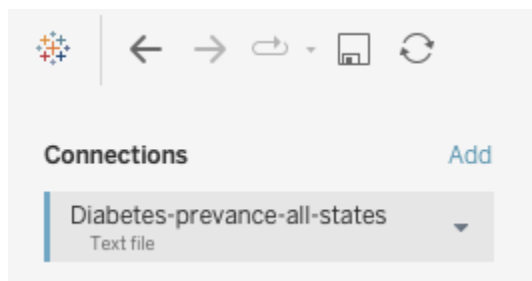
## Datasets:

- Diabetes-prevalence-all-states.csv
- Obesity-Prevalence\_Estimates.csv
- USA-states (map folder)

We will create several graphs and a dashboard for analysis on diabetes incidences across states and overall obesity rates, and answer a few questions regarding the graphs.

## Part 1: Comparison of Diabetes Rates by States

1. Connect to the dataset *Diabetes-prevalence-all-states.csv*

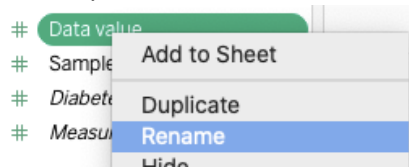


2. Hide the following columns to simplify your table:

*Locationabbr, Class, Topic, Break Out, Break Out Category, Display order, Data Value Footnote, Data Value Symbol, Data Source, Class Id, Topic Id, Breakout Id, Break Out Category Id, Question ID, and Response ID.*

3. Click on a new sheet to start your visualization.

Rename your field "Data Value" to "Diabetes % Rate".

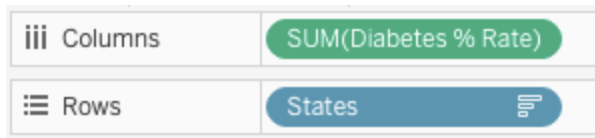


Rename "Locationdesc" to "States".

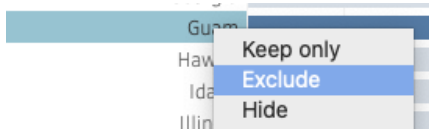
#	Abc	Abc	Abc	#	#	#	#
Diabetes-pr...	Diabetes-prevalence-all...	Diabetes-prevalence-all-states.csv	Diabetes-preval...	Diabetes-prevalence-all...	Diabetes-prevalence-all-s...	Diabetes-prevalence-all-states.csv	Diabetes-prevalence-all-states.csv
Year	States	Question	Response	Sample Size	Diabetes % Rate	Confidence limit Low	Confidence limit High
2022	California	Have you ever been told by a d...	Yes	1,220	11.5000	10.6000	12.4000

4. We will create a bar chart that represents the diabetes rates by State excluding *All States and DC Median, All States, DC and Territory Median, Guam, the Virgin Island, and Puerto Rico*.

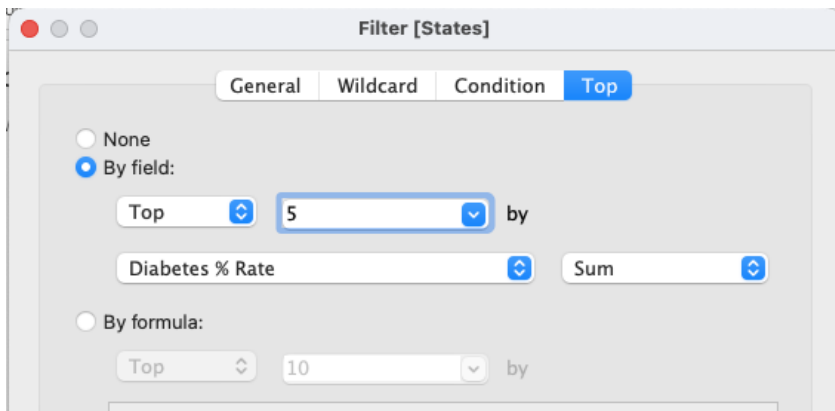
5. Drag *States* to Row and *Diabetes % Rate* to Column



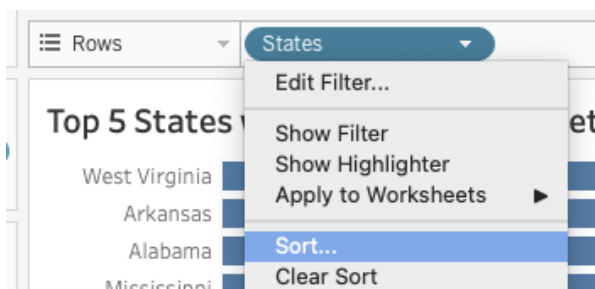
Right-click > Exclude on the bar for values of *All States and DC Median*, *All States, DC and Territory Median*, *Guam*, *the Virgin Island*, and *Puerto Rico*.

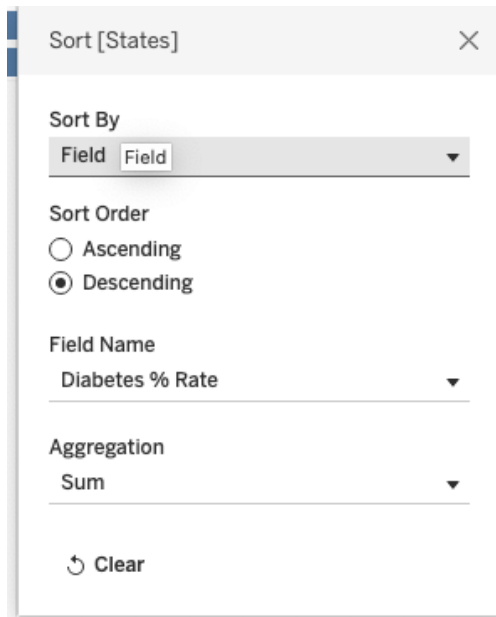


6. Drag *States* to Filter as follows, to display only the 5 States with the highest rates.

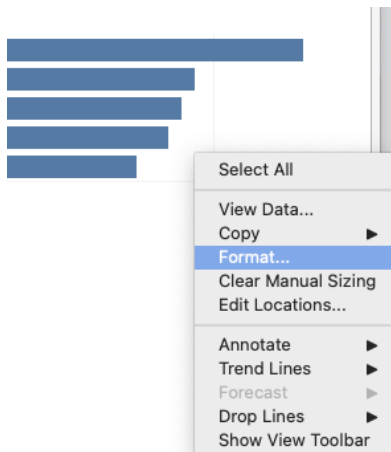


7. Click on the *State* pill in the Row shelf and sort your bars by descending rate order as follows.

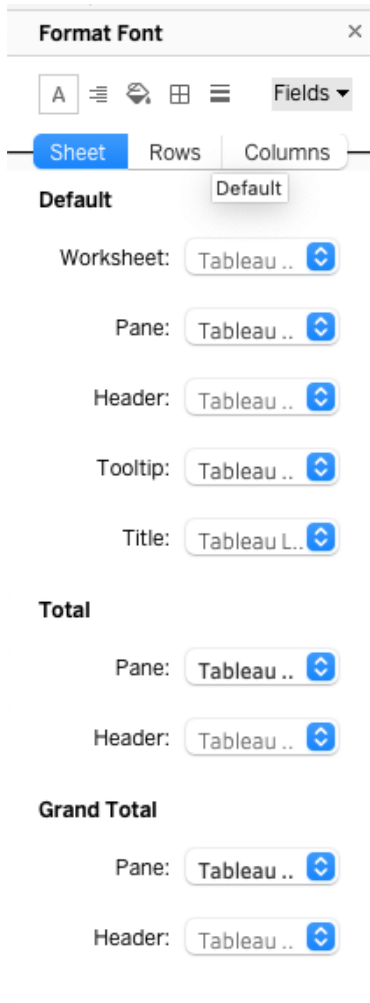




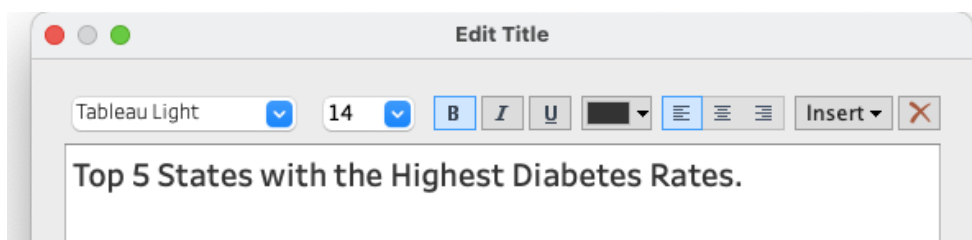
8. Remove all noise from the graph by using right-click > Format on the graph area.



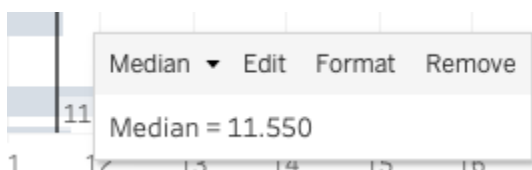
This brings up the Formatting panel on the left side for any formatting.



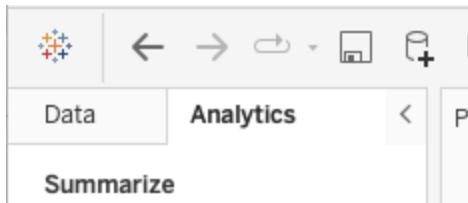
Double-click the title to add a title to the sheet and tab, and format it as follows:



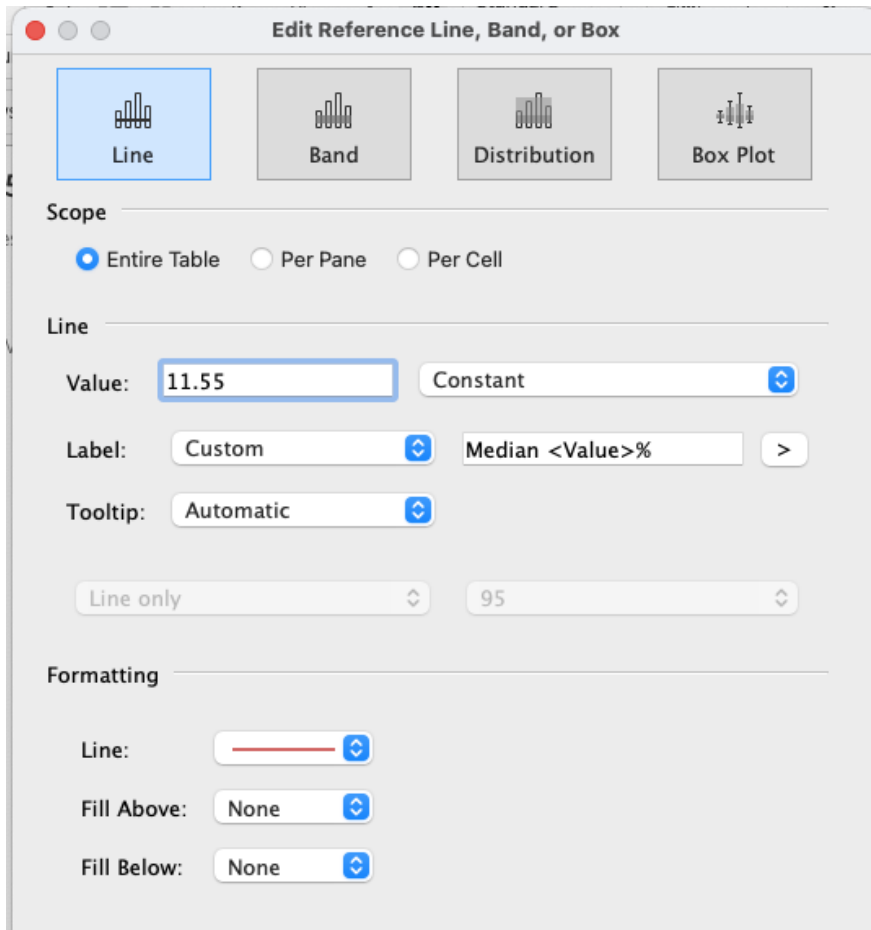
9. We will add a reference line representing the MEDIAN for all States.



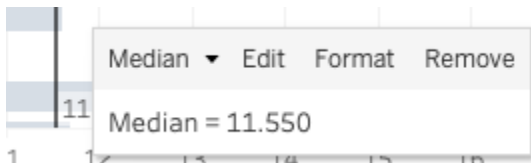
Click on the Analysis tab on the left, next to the Data Panel



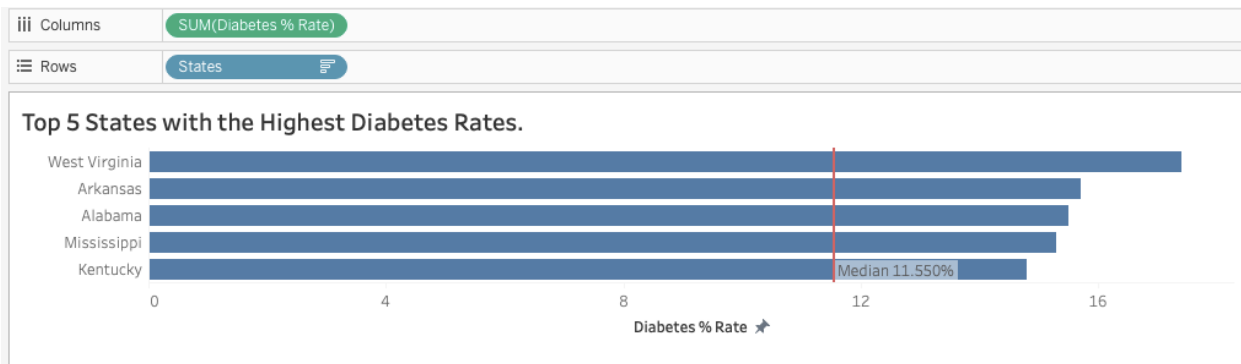
10. Drag a Reference line on to the graph and use the following settings:



Note: The value for the Median was previously determined using a simple exploration graph with all the States in the view.

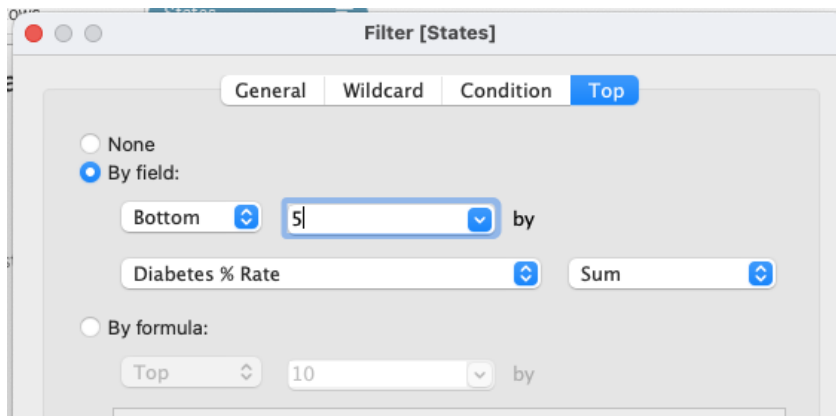


Your final graph should be as follows:

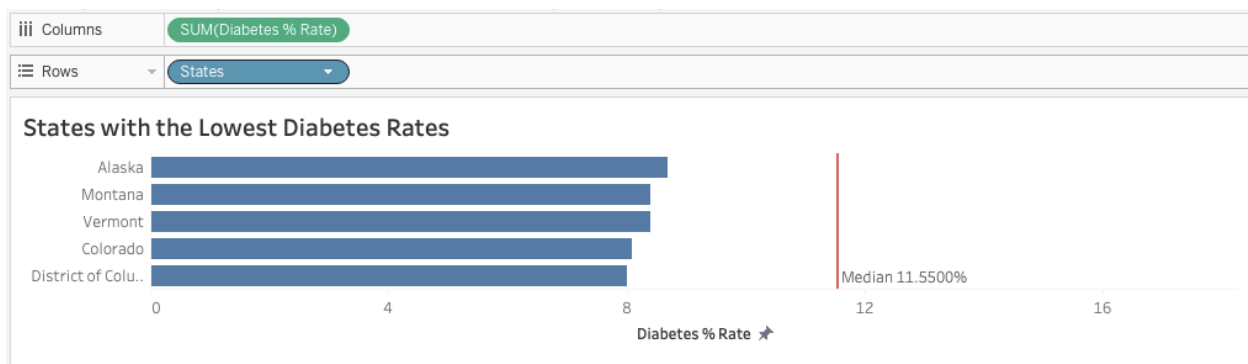


- Using similar steps, create a similar graph, with similar formatting, and decluttering, showing the 5 States with the lowest rate of diabetes.

In the Filter, will we use “Bottom” values.



- Name your sheet and tab “States with the Lowest Diabetes Rates”.

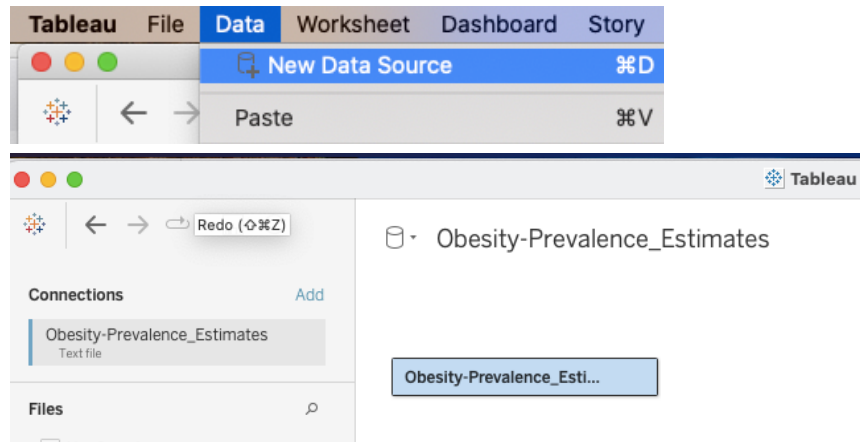


### Analysis questions:

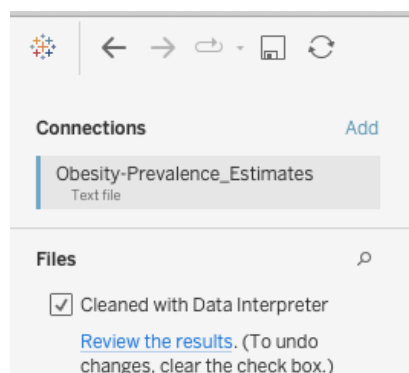
What is the main insight from graph (1), and from graph (2). Analyzing both visualizations together, what additional insight can you draw for them?

## Part 2: Comparison of Obesity Rates by Gender Over Time

1. Click on a new sheet to start a visualization.
2. Connect this sheet to the data source *Obesity-Prevalence\_Estimates.csv*



3. Click on the Data Interpreter checkbox to clear any issue with the dataset.



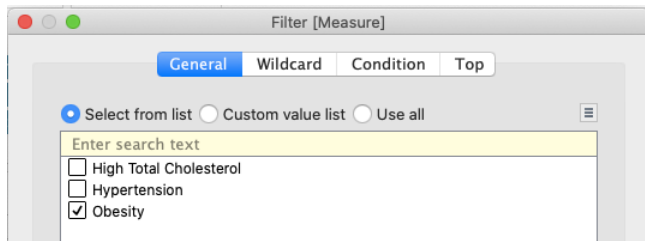
4. In your table, rename your column "Percent" to "Obesity Rate (%)".

a.c...	# Obesity-Prevalence_Estima.c...	# Obesity-
	Obesity Rate (%)	Stand:
	30.3000	
	27.1000	

5. Create a graph showing the obesity rate for the various years of data.  
You will need to do the following:

- a) Drag *Survey Years* to Column and *Obesity Rates* to Rows
- b) Filter by *Age Group* and show only *20 and Over* values, and display the filter ("Show filter" option).
- c) Filter by *Race and Hispanic Origin* and show *All Values*.

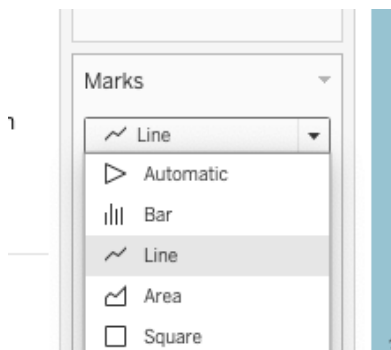
d) Filter the dimension named *Measure* and show only *Obesity*.



e) Drag *Sex* dimension to both details and to color to deaggregate the values. You should get the following stacked bar chart.



We need to modify the graph to present it as a line chart. Click on mark and select line from the drop-down menu.





6. Edit your colors to make them more distinct from each other. After some additional formatting and decluttering, your graph should look like the following:



### Analysis Questions:

- How can you modify your graph to analyze all three values within the *Measure* field (*Obesity*, *High Total Cholesterol* and *Hypertension*)? Note this is a categorical field, not to be confused with the Measure in Tableau.
- What insights can you draw from the graph? Give three, in order of importance.

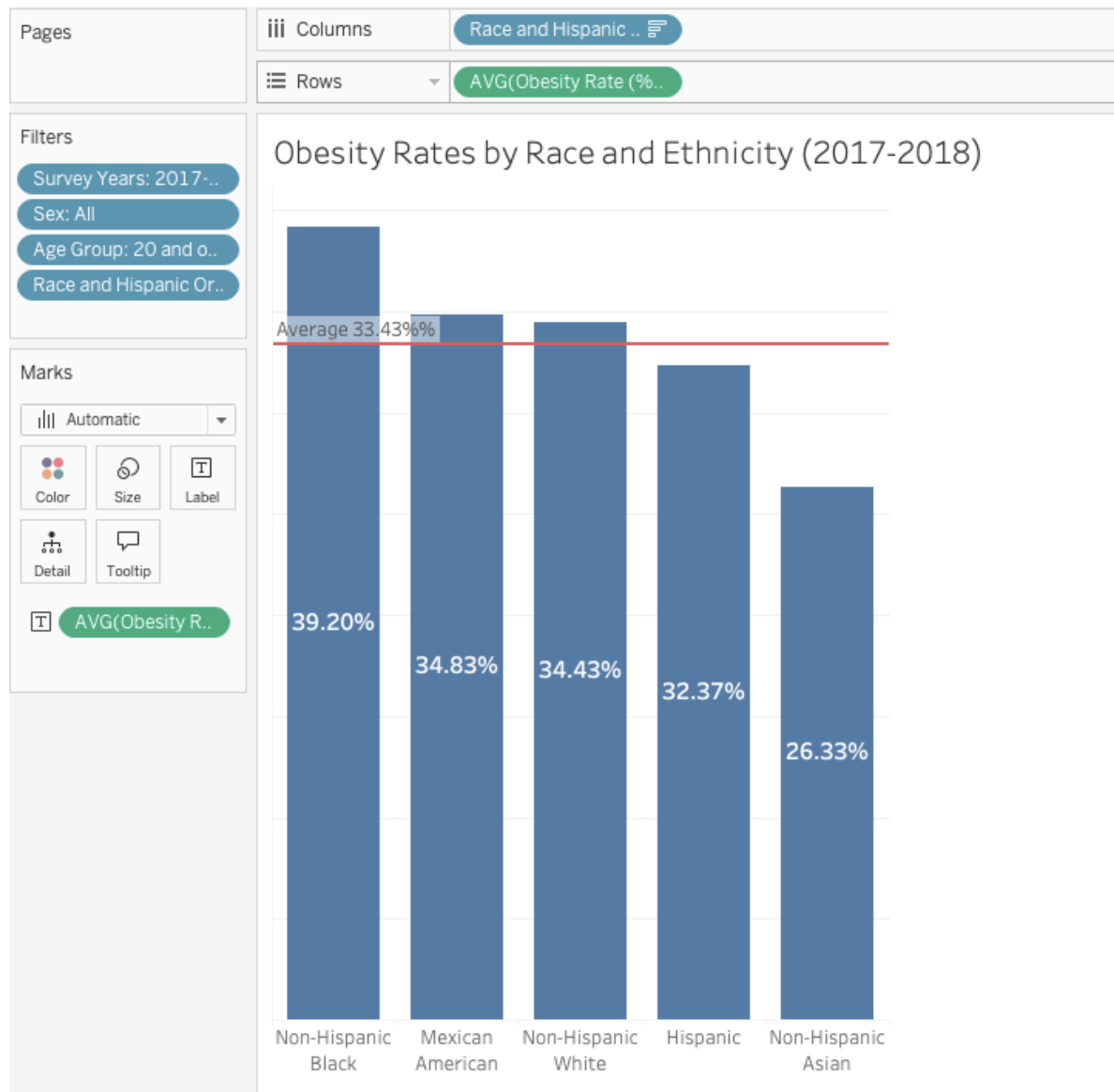
### Part 3: Comparison of Obesity Rates by Race and Ethnicity

- Using similar filters as in Part 1, create a bar chart showing the average obesity rate by race as follows:
  - Survey Year: 2017-2018
  - Sex: All
  - Age Group: 20 and over
  - Measure Names: Obesity Rate %
  - Race and Hispanic Origin: Exclude "All"

Note: Your measure should be aggregated as an average (AVG).

- Add a reference line with the AVG Obesity Rate. Format it with the value to show "33.4% Average Obesity Rate".

After some formatting, your graph should be as follow:



**Analysis question:** What insight can you draw from the graph?

#### Part 4: Dashboard Analysis

Combine all your graphs into a Dashboard, size: 1028 x768.

Attention should be given to:

- Layout and structure, alignment and positioning of the graphs and other elements
- Removing noise from the graphs
- Incorporate interactive features like filters for analysis.
- Guiding users through the data with storytelling elements and insights.