DSCI 5360 – Data Visualization
Assignment 1

Instructions

1. This assignment is to be submitted online through Canvas, on or before the due date.
2. This assignment must be completed using the version of Tableau specified in the syllabus. I will not grade assignments completed with newer or older versions of the software.
3. Please name your file "Assignment1_FirstName_LastName.twbx" or "Assignment1_Group_GroupNumber" as appropriate. Please note the ".twbx" extension. I will be unable to grade files with the ".twb" extension.
4. Your Tableau workbook should have separate worksheets clearly labeled as "Question 1," "Question 2," … etc.
5. If a question asks you to do more than simply create a visualization (i.e., answer a specific question in text), please do so in the title of the worksheet or in a caption/annotation.
6. When asked to recreate a visualization, you should attempt to replicate it exactly in terms of data, color, marks, annotations, etc. The exception is axis labels, which are dynamically altered by Tableau based on the display size of the screen.
7. This assignment is to be completed without external assistance. You are welcome to have general discussions with your classmates about Tableau functionality, but there should be no specific reference to the assignment.

Datasets

- movie_metadata.xlsx
- GDP_2018.csv
- GDP_2019.csv

Case 1: The Movies

The Excel file "movie_metadata.xlsx" contains IMDB information for approximately 5000 movies. Follow the steps below and create Tableau worksheets to answer each of the following questions.

1. **Recreate the visualization shown in Figure 1 that depicts the top grossing movie directors. Who is the eighth top grossing movie director?**

   HINTS:
   a. Notice that the data is sorted in descending order based on the sum of movie gross by director.
   b. The label for Gross has been formatted to show values as currency in Billions. Right click>format>fields>SUM(Gross)>Pane>Default>Numbers>Currency (Custom)

2. **Recreate the visualization shown in Figure 2 that depicts G- and R-rated movies by Gross and IMDB score, where Gross is at least $200,000,000. What movie is annotated in the figure?**

   HINTS:
   a. Start by adding SUM(Imdb Score) and SUM(Gross) to your columns and rows respectively
   b. The drag Movie Title and drop it on your visualization. This should create the initial scatter plot.
   c. Now you need to filter out those movies with less than $200M gross and those with ratings other than G or R. Drag these fields to the filter shelf and configure appropriately. If asked, choose "All Values."
   d. Change marks from "Automatic" to get the appropriate marker.
   e. Notice the movies are color coded by content rating. You will need to drag that field to the "Color" card, then double click on the color legend to set the appropriate colors.

3. We would like to see which European country produces the highest grossing movies. **Recreate the visualization in Figure 3 and make sure to apply annotations to the top 3 countries.**

   HINTS:
   a. Notice that the country dimension has a geographic role. Double clicking that dimension will automatically create a map with a marker for each country in the dataset
   b. Dragging SUM(Gross) to the "Size" card will size the markers based on the sum of gross for all movies produced in each country.
   c. You can increase or decrease the size of the markers by clicking on the "Size" card and adjusting the slider.
   d. Since we are interested in Europe, you should zoom in on that section of the map and adjust the size appropriately.

Case 2: GDP

We are interested in differences in GDP in the years 2018 and 2019. Using the csv files "GDP_2018.csv" and "GDP_2019.csv," follow the steps below and create Tableau worksheets (In the same Tableau workbook you used above) to answer each of the following questions.

HINT: To add the new data sources, create a new sheet and then select Data>New Data Source. Click "More" under To a File. Navigate to GDP_2018.csv. Examine the field names and notice that they come in as "F1," "F2," etc. Tableau hasn't recognized that our first row are headers. Click the "Use Data Interpreter" checkbox to resolve the issue. Depending on your environment, you may see both the GDP_2018.csv and GDP_2019.csv files displayed. If not, click "Add" next to Connections, click "More" under To a File. Navigate to GDP_2019.csv. Once you see both GDP_2018.csv and GDP_2019.csv, you can set up a join. Double click the gray GDP_2018.csv box on the canvas, Tableau should indicate that GDP_2018.csv is made of 1 table. Drag GDP_2019.csv to the canvas. You should see a join between the two files. Click the join and make sure it is specified as "Inner"

and that the files are joined on Country Code. Close out of the data source and move back to your sheet. You should see the dimensions and measures associated with both files.

4. **Please recreate the visualization shown in Figure 4 in which the color in the top map describes the SUM(2018) GDP and the color in the bottom map describes the SUM(2019) GDP.**

   HINTS:

   a. Notice that the GDP_2018.csv country code dimension has a geographic role. Double clicking that dimension will automatically create a map with a marker for each country in the dataset
   b. Dragging another instance of Latitude (generated) to the Rows shelf will create the second map. Notice that under marks you now have three sections. All refers to both maps, the first Latitude (generated) section refers to just the top map and the second refers to the bottom map. Make sure you are making changes to the appropriate section.
   c. You want to color your top map based on the SUM(2018) GDP. Click the appropriate section under marks and drag the field to the color card. Then do the same for the 2019 map.
   d. Change colors as appropriate by clicking on the color legend associated with the map you want to alter.
   e. Add area annotations to designate what each map represents.

5. **To understand year over year changes, please recreate the view shown in Figure 5.**

   HINTS:

   a. Start by moving Country Code to the rows shelf then drag 2018 to the "ABC" column in your visualization.
   b. Now drag 2019 and drop it on the 2018 column. Several changes happen at this point because we are essentially trying to represent multiple pieces of information in the same space. Tableau recognizes this and substitutes special "Measure Names" and "Measure Values" variables.
   c. You need a calculated field to capture the difference between 2019 and 2018. Click Analysis>Create Calculated Field. Name it "GDP_Diff" and use the formula below:

      ```
      [2019] - [2018]
      ```

   d. GDP_Diff should be added as a new measure. Drag it to your visualization.
   e. You need a filter to exclude data with null values for GDP_Diff. You can find it under the "Special" category when you add the filter.
   f. The values are color coded based on whether the GDP_Diff is positive or negative. This will require another calculated field which will then be added to the "Color" card.

Figures

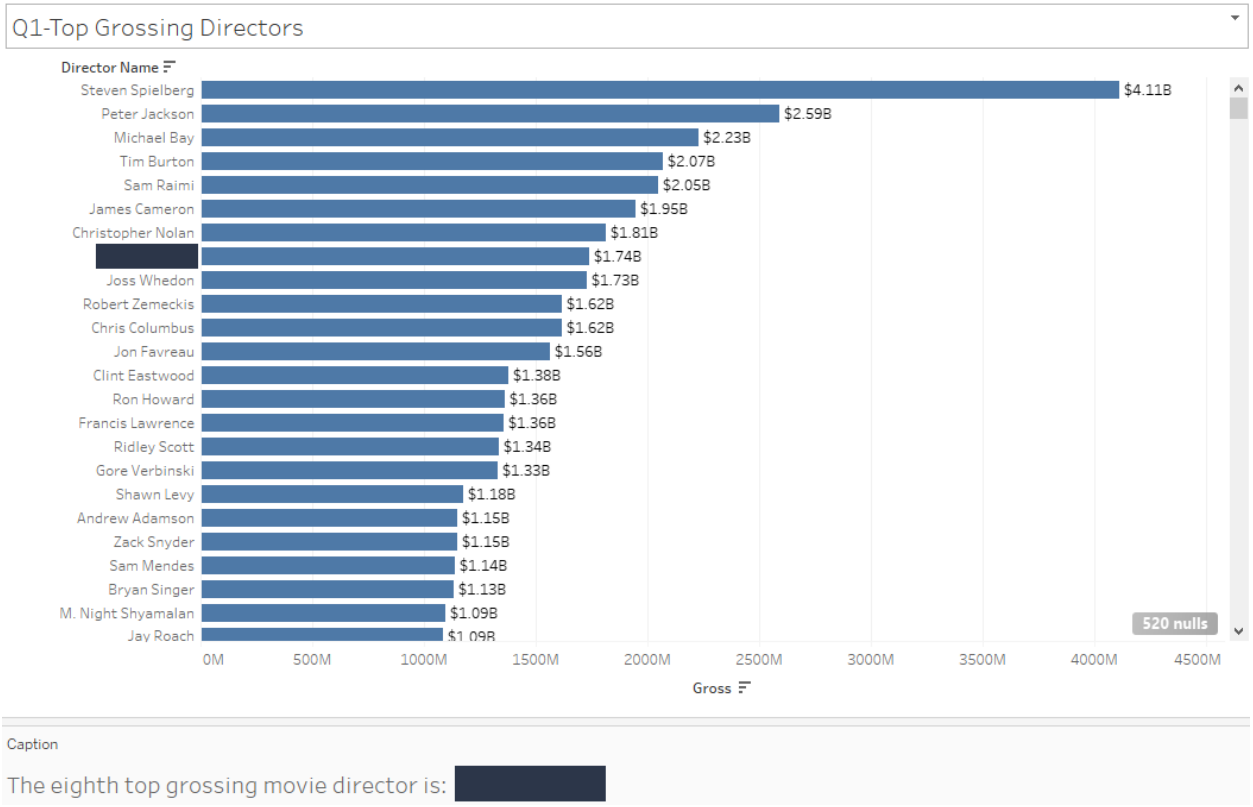

Figure 1

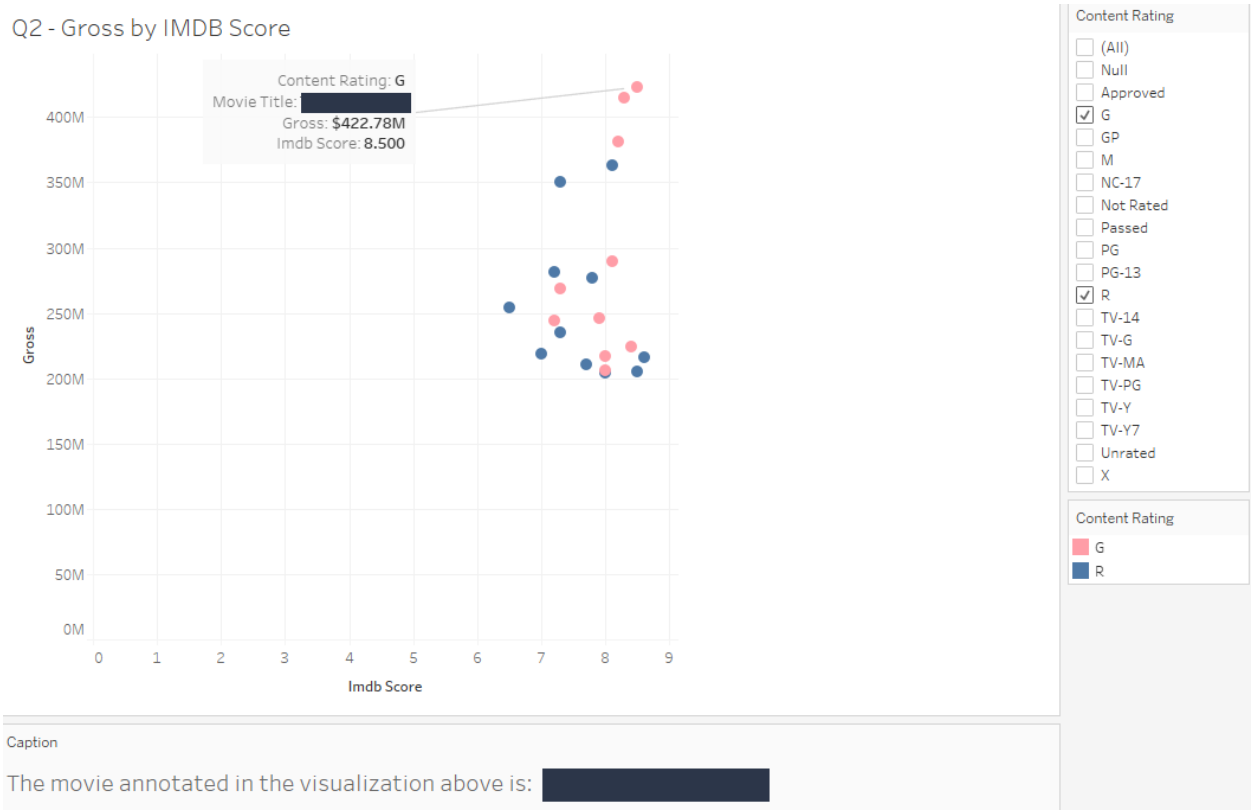
Q2 - Gross by IMDB Score
Content Rating: G
Movie Title: [redacted]
Gross: $422.78M
Imdb Score: 8.500
400M
350M
300M
250M
200M
150M
100M
50M
0M
Gross
0
1
2
3
4
5
6
7
8
9
Imdb Score
Content Rating
(All)
Null
Approved
G
GP
M
NC-17
Not Rated
Passed
PG
PG-13
R
TV-14
TV-G
TV-MA
TV-PG
TV-Y
TV-Y7
Unrated
X
Content Rating
G
R
Caption
The movie annotated in the visualization above is: [redacted]

Figure 2

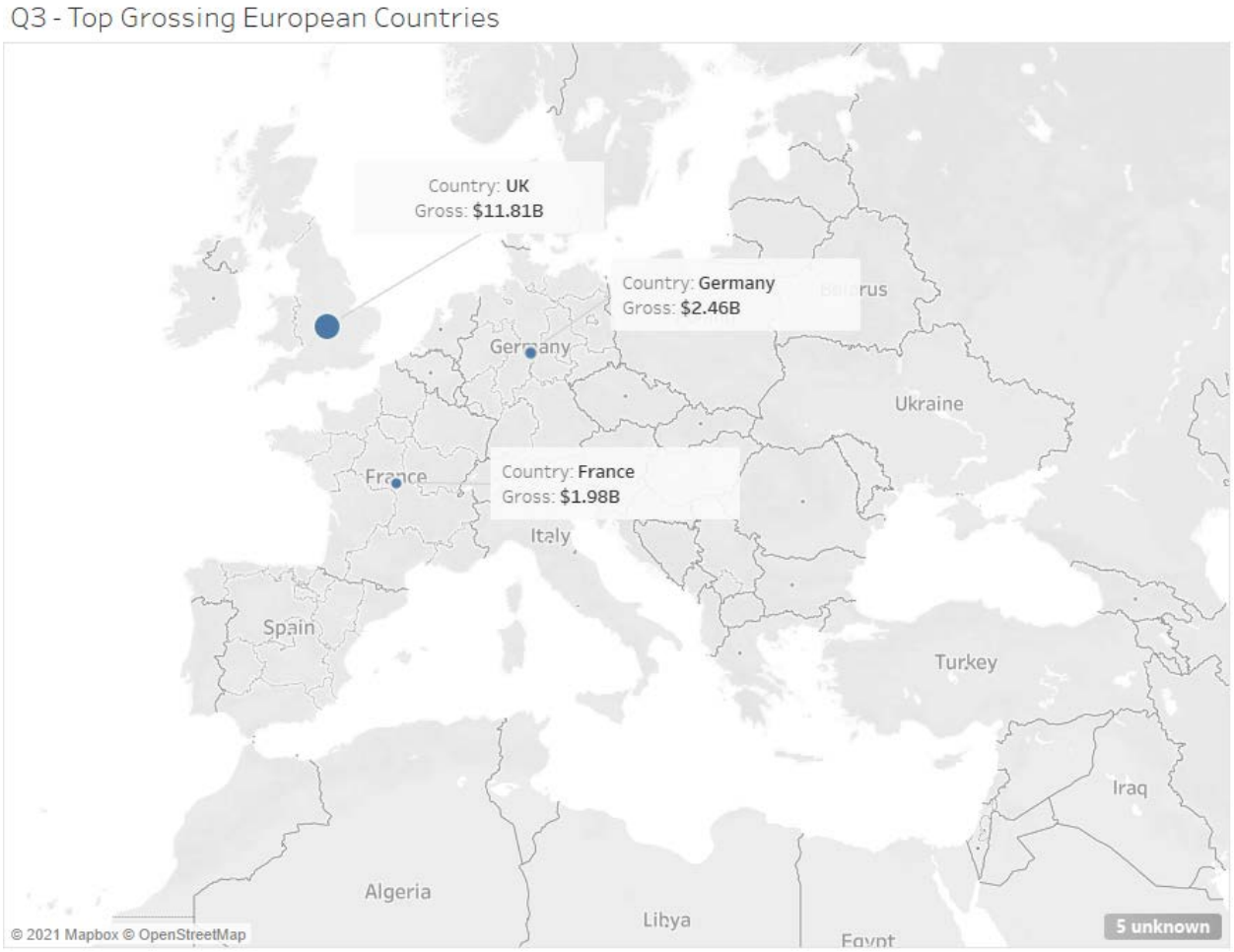
Q3 - Top Grossing European Countries
Country: UK
Gross: $11.81B
Country: Germany
Gross: $2.46B
Germany
Ukraine
France
Country: France
Gross: $1.98B
Italy
Spain
Turkey
Iraq
Algeria
Libya
Egypt
© 2021 Mapbox © OpenStreetMap
5 unknown

Figure 3

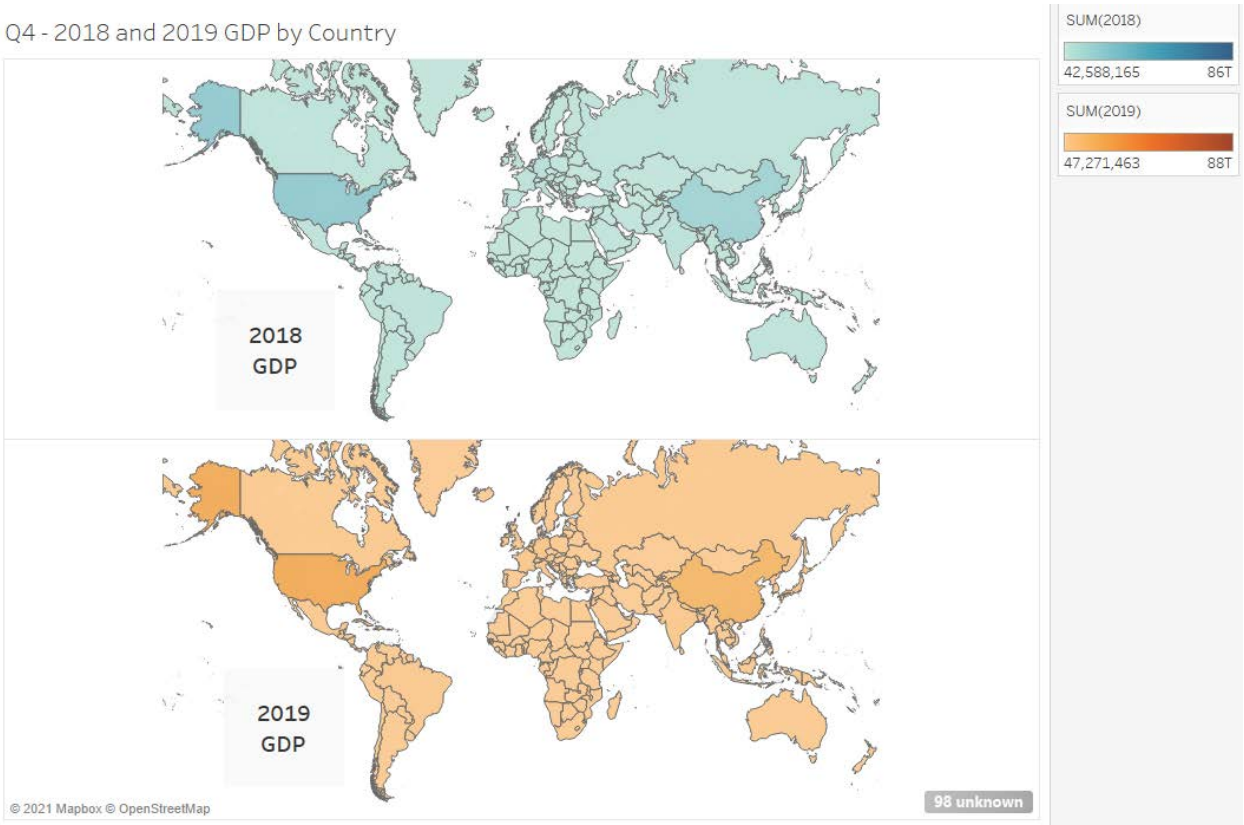
Q4 - 2018 and 2019 GDP by Country
2018
GDP
2019
GDP
© 2021 Mapbox © OpenStreetMap
98 unknown
SUM(2018)
42,588,165
86T
SUM(2019)
47,271,463
88T

Figure 4

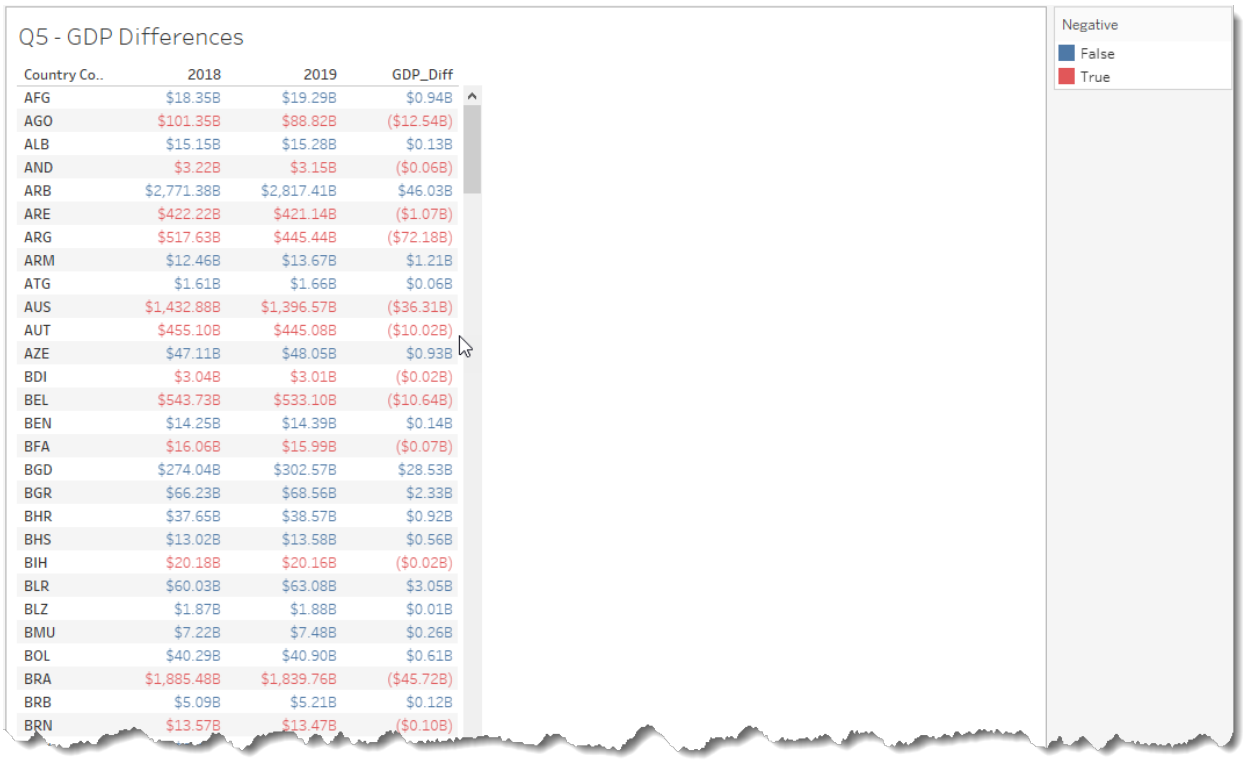
Q5 - GDP Differences
Country Co..
2018
2019
GDP_Diff
AFG $18.35B $19.29B $0.94B
AGO $101.35B $88.82B ($12.54B)
ALB $15.15B $15.28B $0.13B
AND $3.22B $3.15B ($0.06B)
ARB $2,771.38B $2,817.41B $46.03B
ARE $422.22B $421.14B ($1.07B)
ARG $517.63B $445.44B ($72.18B)
ARM $12.46B $13.67B $1.21B
ATG $1.61B $1.66B $0.06B
AUS $1,432.88B $1,396.57B ($36.31B)
AUT $455.10B $445.08B ($10.02B)
AZE $47.11B $48.05B $0.93B
BDI $3.04B $3.01B ($0.02B)
BEL $543.73B $533.10B ($10.64B)
BEN $14.25B $14.39B $0.14B
BFA $16.06B $15.99B ($0.07B)
BGD $274.04B $302.57B $28.53B
BGR $66.23B $68.56B $2.33B
BHR $37.65B $38.57B $0.92B
BHS $13.02B $13.58B $0.56B
BIH $20.18B $20.16B ($0.02B)
BLR $60.03B $63.08B $3.05B
BLZ $1.87B $1.88B $0.01B
BMU $7.22B $7.48B $0.26B
BOL $40.29B $40.90B $0.61B
BRA $1,885.48B $1,839.76B ($45.72B)
BRB $5.09B $5.21B $0.12B
BRN $13.57B $13.47B ($0.10B)
Negative
False
True

Figure 5