

Lab 1 Handout

DUE: Tuesday Sept. 6 at 9:00 PM

Get Tableau now!

Please make sure you have Tableau Desktop installed and running. You can get a copy at:
<https://www.tableau.com/products/desktop/download>

Product key: TCR7-B43B-E010-0B40-E915

TL;DR

What you'll need (download from Canvas):

1. Online copy of this document (**week1_tableau_handout.pdf** in canvas)
2. Movie dataset (**movies_independent_part_1.xlsx** in canvas)
3. Bechdel dataset (**movies_independent_part_2.csv** in canvas)
4. Answer sheet for part 1 (**independent_part_1_answer_sheet.docx** in canvas),
5. Answer sheet for part 2 (**independent_part_2_answer_sheet.docx** in canvas)

What to turn in: Four files (2 x PDF and 2 x Tableau) –

1. Answer sheet (PDF) for part 1 of the independent activity
2. Tableau workbook for part 1 of the independent activity (you do not need to replicate everything we went over during the lab, just answer the questions that are listed in the answer sheet)
3. Answer sheet (PDF) for part 2 of the independent activity
4. Tableau workbook for part 2 of the independent activity

When to turn it in: Tuesday Sept. 6 at 9 PM. All technical labs are due Tuesday (see the schedule on canvas). See the lateness policy in the syllabus for more information (one 48 hour freebie, and penalties for the rest). Labs build off each other, therefore, keeping up is important. Do your best with these. We know some of you are building up coding skills at the same time as learning the APIs. We *do* give partial credit so turn in what you have.

General guidelines for our technical labs:

- You can collaborate on the *lab activity* with others, but please let us know who you worked with. For the second part, the Bechdel article, please work on your own. It's ok to talk through general strategies with others, but this should be your own work.
- In general, the focus of these labs is to get you familiar with visualization tools. We will provide you with **a curated tutorial, a dataset, and a task**. We will give you a brief demo/intro about the topic and task at the beginning of the lab.
- We've tried to make this class a great experience for all students, regardless of background. You might be more or less familiar with the tool we are teaching. **Don't panic if you feel lost during some tutorials, we are here to help you.** If you run into roadblocks, please don't hesitate to ask. Here are some (but not all) ways to get help
 - Google it! There are lots of resources out there. We'll also post resources based on the topic.

- Talk to other students (just don't copy!).
- You can email us, send us a slack message, post on a slack #channel, or come to office hours. We have various forums to ask for help.

Lab 1 Specifics

Follow along with our demonstration video on Perusall (under Lab videos, called "Tableau 101"). We will be demonstrating many parts of this lab. As you watch our demonstration, make sure you replicate our steps. We've tried to make this document as detailed as possible so you can use it outside of class/videos.

In this document you will find some red squares. **Red squares □ indicate things you should do on your own. Record your answers in your answer sheets.**

Tableau is used to generate interactive data visualizations and makes very common visualizations very easy. Our focus today is on creating simple visualizations using Tableau.

In this lab you will learn:

Data	Know how to import data
	View and describe data within Tableau
	Sort data
	Select data types (Date, Number, String)
	Make Calculated Fields
	Change the # of rows displayed
	Create Groups
Sheet	Differentiate measure and dimensions
	Create simple visualization and change chart type
	Adjust encodings
	Adjust color
	Adjust aggregation methods (e.g. sum, average, count)
	Sort axis
Organize Sheets	Build dashboard
	Build stories
Action	Create annotations on the chart
	Create data filters
	Link sheets together with actions.
	Add analytics enhancements (e.g., constant line, regression)
	Edit legend and filter by legend.

Tableau Basics

Make sure you can run Tableau (see first point above). If you need more information on Tableau take a look at:

- Videos: <http://www.tableau.com/learn/training>
- Books (for both these you have access through the library):
 - Practical Tableau: <https://bit.ly/3kVvE0j>
 - Jumpstart Tableau: <https://bit.ly/3kZaeze>
 - There are a number of other videos and books on the O'Reilly site
- Cheat sheets (these are a bit old, but still good):
 - <https://bit.ly/3h9w2Gh>
 - <https://bit.ly/3h8TSlT>

GLOSSARY OF TERMS:

Sheet: A sheet is a singular chart or map in Tableau.

Dashboard: A dashboard is a canvas for displaying multiple sheets at a time and allowing them to interact with each other.

Workbook: A workbook is the entire Tableau file containing your sheets and dashboards.

Measure: A variable from the dataset that is meant to be aggregated. (This means it should be a number that it makes sense to do math with: sum, average, and so on.) Measures are often continuous data. Examples include GPA, sales, quantity, quota, height, and salary.

Dimension: A categorical variable from the dataset that is used to slice and dice the data into different categories. Dimensions are often discrete data. Examples include country, gender, student ID, and name.

Filter: A filter is used to limit what data is being displayed on the sheet. Visible controls for a filter on a sheet or dashboard are called Quick Filters.

Tooltip: Tooltips are text boxes that appear when hovering over a mark on a sheet in order to give more information. The text and text formatting in them are easily edited through the Marks card.

Marks card: The Marks card is the tool used to create a sheet that controls most of the visual elements in a sheet. Using the Marks card, you can switch between different chart types (bar, line, symbol, filled map, and so), change colors and sizes, add labels, change the level of detail, and edit the tool tips.

Rows and Columns Shelves: The Rows shelf and the Columns shelf is where you determine which variables will go on what axis. Put data you want displayed along the X-axis on the Columns shelf and data you want displayed on the Y-axis on the Rows shelf.

Guided Demo – The Movie Dataset

Scenario: You work for a company that is getting into the movie business. You have been hired to help decide which future movies your company should invest in. Your bosses have lots of questions: Which directors are good? What genres? When should movies be released? They would like to see a report for their various questions.

PART 1: Importing and basic data manipulation

- 1) Download the movies dataset from canvas (under 01 - Intro/lab) open the Tableau app and load the data.

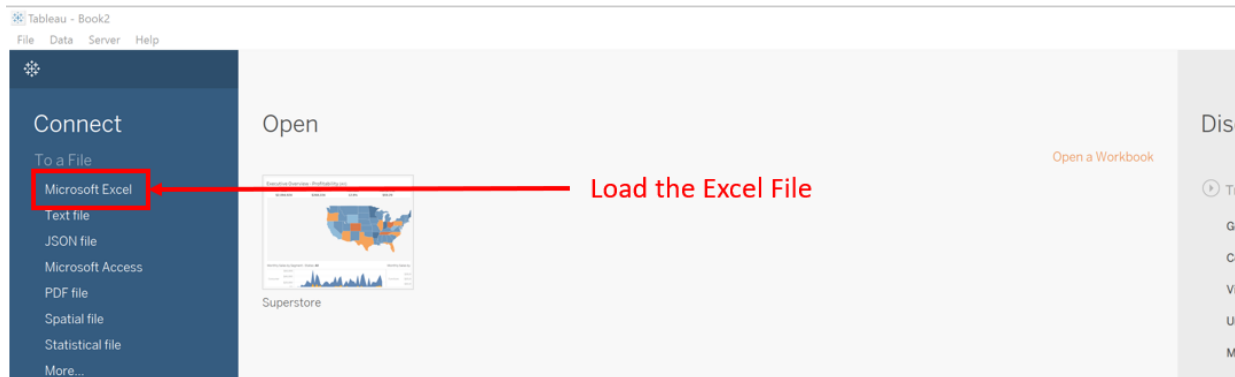


Figure 1

- 2) Once you load it, you will see something like this:

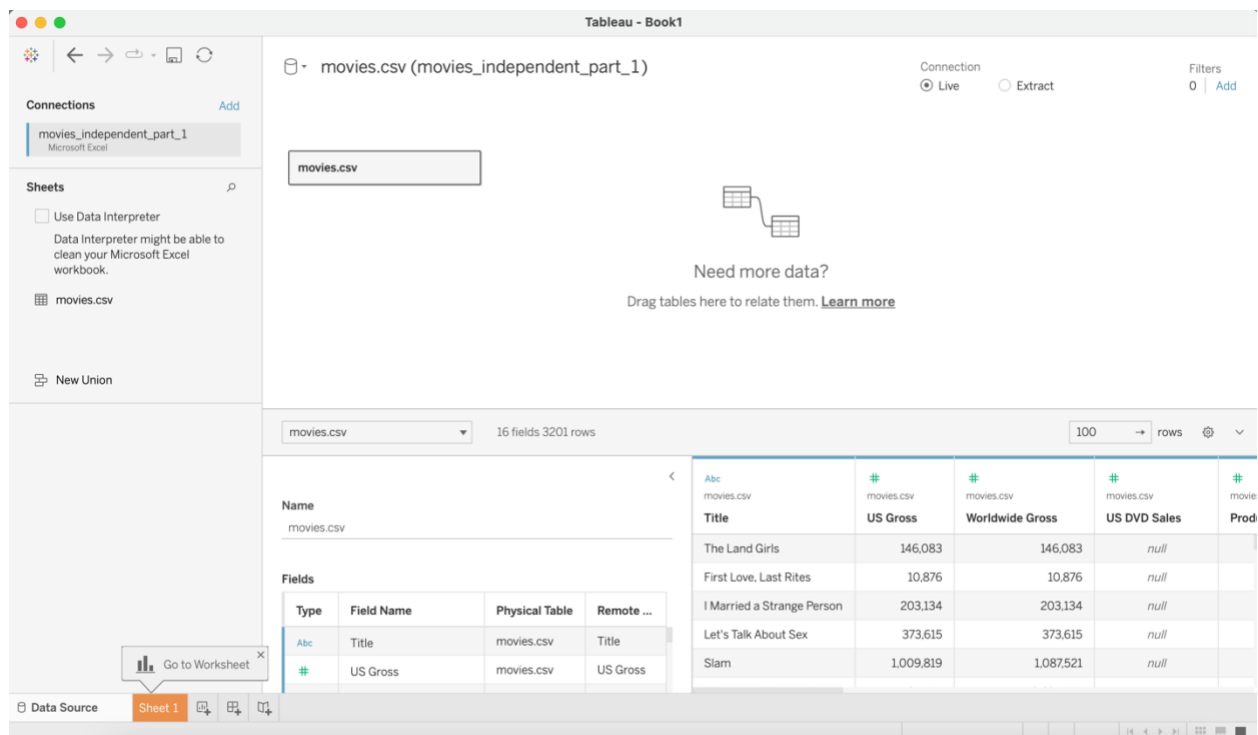
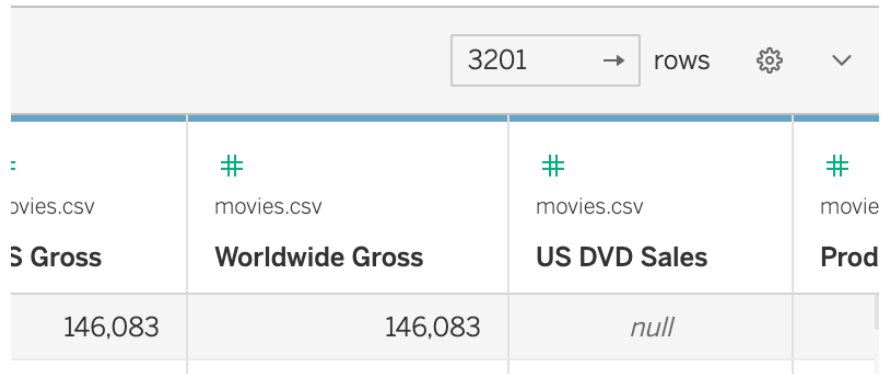


Figure 2

You can view and edit the data under the “Data Source” Tab.

- 3) Change the number of rows shown to the maximum rows (3201). Make sure that you are displaying *all* rows in the data, or your results might look a little different from ours in this tutorial. (Figure 3)

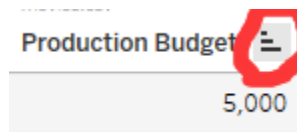


3201 → rows ⚙️ ▼						
movies.csv	#	movies.csv	#	movies.csv	#	movie
US Gross	Worldwide Gross	US DVD Sales	Prod			
146,083	146,083	null				

Figure 3

Question 1: Which movie has the highest production budget? Which movie has the lowest production budget?

- 1) To answer this, we'll simply sort the dataset by production budget (Figure Q1.1). No fancy visualization required!



Production Budget
5,000

Figure Q1.1

Question 2: What is the range of release dates?

- 1) Sort the table by "Release Date".

Question 3: Which is the most profitable movie in our dataset?

Unfortunately, we don't have a column for this (revenue for us will be "gross" minus "production cost"). Good news: it's easy to create!

- 1) Sort the dataset by the "worldwide gross"
- 2) Create a **calculated field** called "Revenue" that shows the actual profit that these movies make (hint: Revenue= Worldwide Gross-production budget, assuming that we only care about movies that target the international market) (Figure Q3.1)

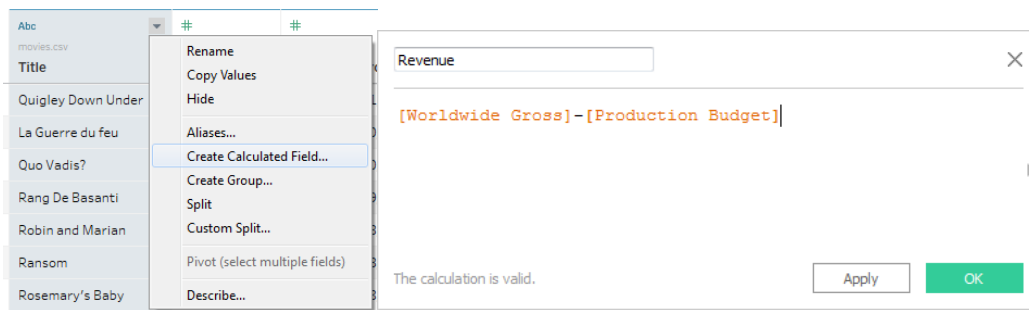


Figure Q3.1

Question 4: How many directors are included in the dataset?

- 1) Click on the triangle next to "Director" -> select "Describe". The "Describe Field" window shows the answer. (Figure Q4.1)

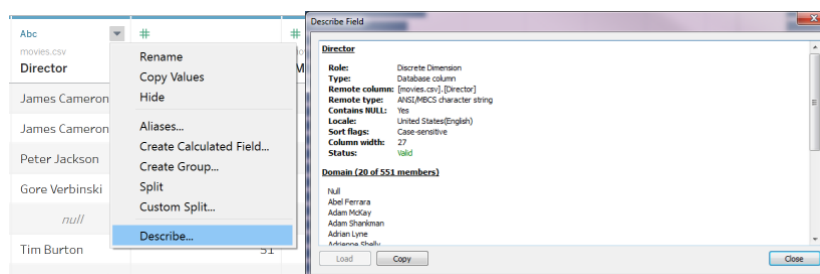


Figure Q4.1

Question 5: I have a hypothesis, movies with certain names have interesting properties. How do I find all movies that contain specific strings in their names?

Here's an example for movies that contain "star":

- 1) Create a group of all movies that contain "star" in their names. (Figure Q5.1)

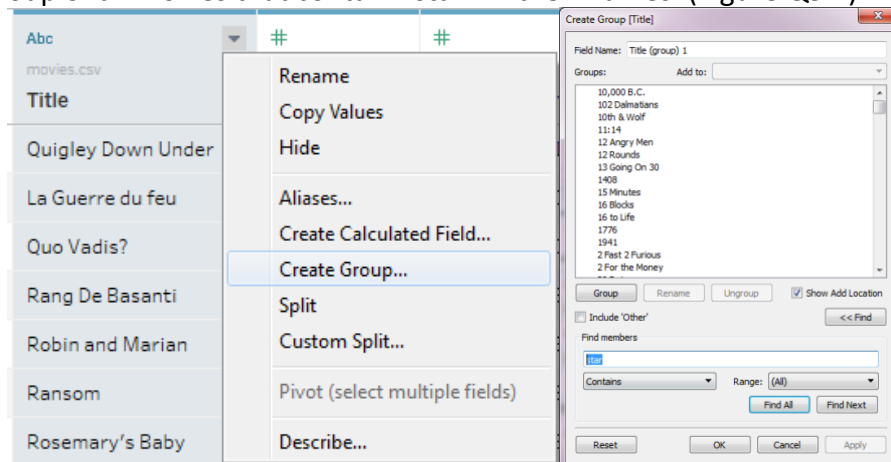


Figure Q5.1

(Summary: Triangle -> create group-> (Click on Find >> to expand) -> enter search criteria-> "Find ALL"-> Click "**Group**" ->Change group name (optional) -> Check "Include Other" (optional) -> Change Field names (optional))

☐ Question 6: Which movies were about dragons?

- 1) Create a group of all movies that contain the string “**dragon**” in their names.
- 2) Paste a screenshot of your display/filter in the answer sheet (we don’t need it all, just a sample)

PART 2: Creating Simple Visualizations with Tableau

Let's try some actual visualizations. Click on the "Sheet1" tab and you will see a layout similar to Figure 4.

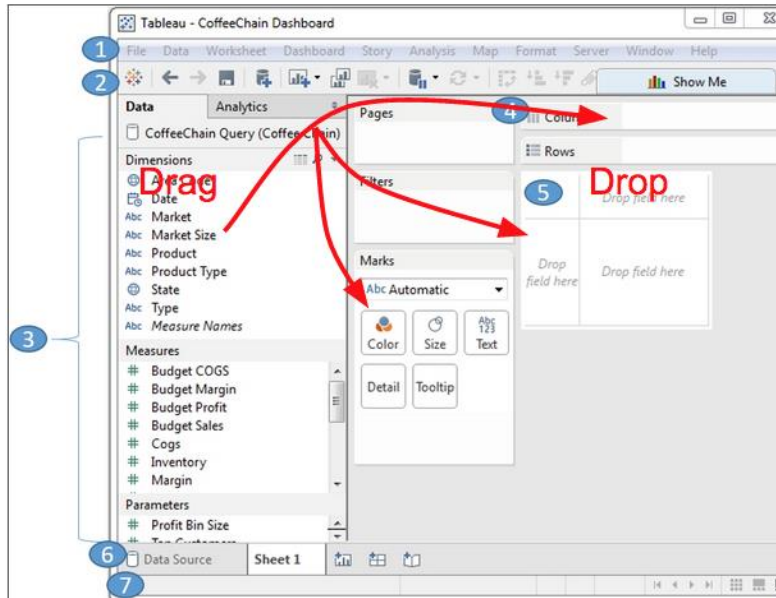


Figure 4

The side bar has two important blocks, **measures** and **dimensions**.

You can think **dimensions** as entities (e.g., each movie) or groups (e.g., all action movies). They will most often manifest column or row headings. Often, you will have a “mark” (e.g., circle, square, etc.) for the dimensions you visualize. **Measures** are exactly what they sound like: something we measure for each dimension (e.g., a specific movie’s revenue, the average production budget for action movies, or the earliest comedy release date). Visually, dimensions will look like little blue pills and the measures will be green. You’ll be able to transform one to the other, but more on that later.

We will see some examples in Questions 7 and 8.

Question 7: Who is the most successful director?

- 1) Drag “Director” to the row (this will be our dimension), drag “Revenue” to the column (this will be our measure). Notice that the measure will automatically change to “sum.” Because a single director can have many movies, Tableau takes a guess about what you want to do with those.

- 2) The “Show Me” menu let you create different types of charts (Figure Q7.1).

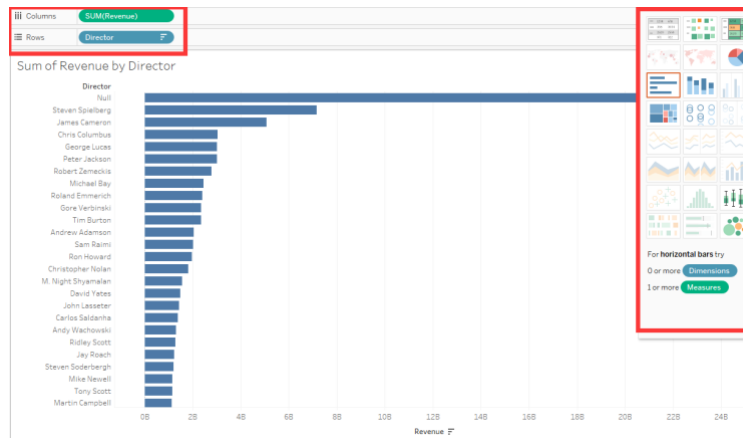


Figure Q7.1

- 3) Filter Null Values in the “Director” field by right-clicking on the “Null” value and excluding the values. (notice that a new filter is created – Figure Q7.2)

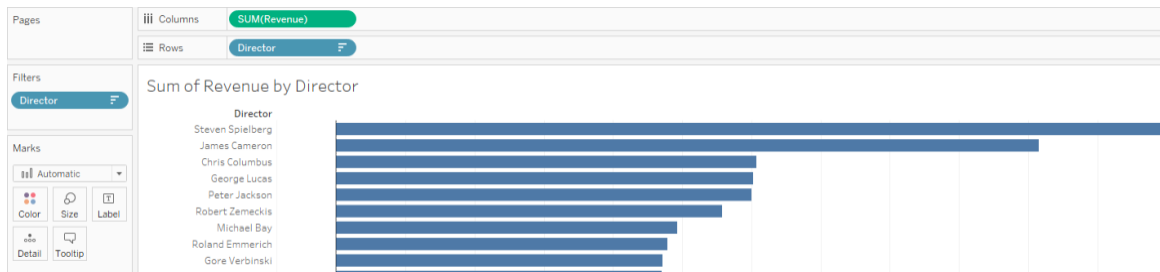


Figure Q7.2

- 4) Drag “Revenue” to column and convert it to the measure of average (Figure Q7.3). Notice Tableau will create the new visualization as a new “column.”

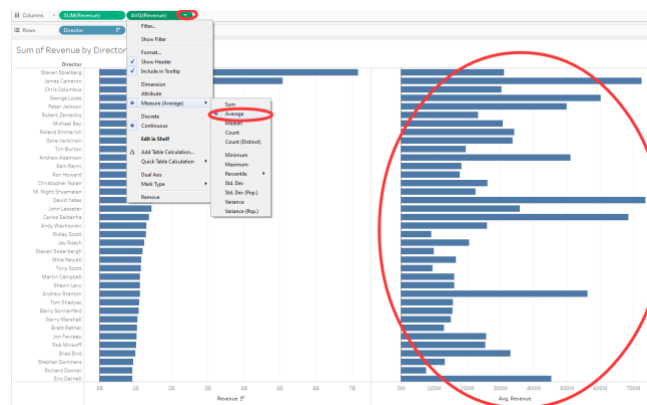


Figure Q7.3

- 5) If you find something interesting (like who is David Yates?!) add annotations (Figure Q7.4)

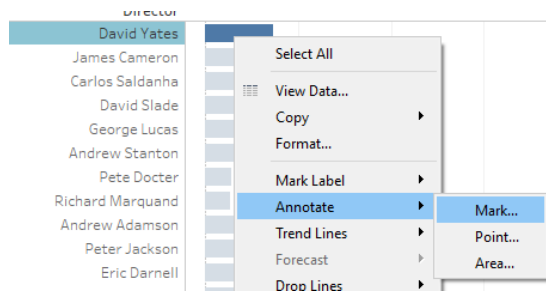


Figure Q7.4. 1: Right click the field/object that you want to annotate

- 6) Edit the color of the bar by clicking on the “Color” option in the “Mark” shelf (we want different colors for the visualizations).
- 7) Edit the Sheet1 title to be “Sum and Average of Revenue by Director”.

Question 8: Are there any directors who have bad profitability?

- 1) Create a new worksheet and name it as “Revenue as Dimension”.
- 2) Drag “Director” to the Rows. Drag “Revenue” to Columns, and convert it to Dimension (Figure Q8.1).



Figure Q8.1

- 3) Add “Title” to the “Mark” shelf as tooltip so that each revenue mark is associated with the movie title (Figure Q8.2)

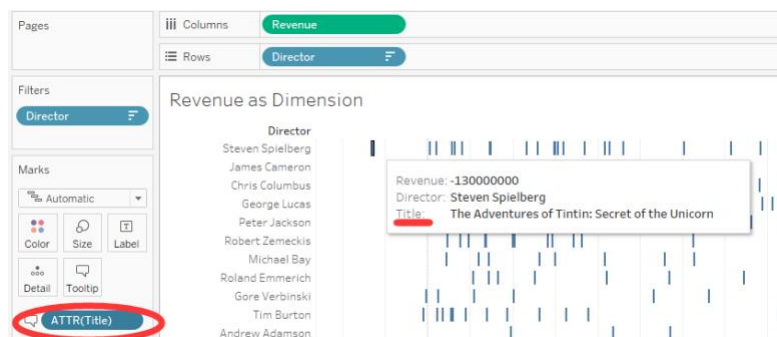


Figure Q8.2

- 4) Add a constant line to indicate movies that have negative revenue (Hint: Analytics-> Summarize->Constant Line-> Value=0) (Figure Q8.3)

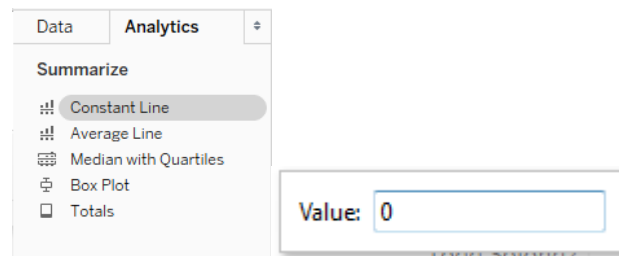


Figure Q8.3

- 5) Drag “Major Genre” to “Marks” shelf and encode it using color. (Figure Q8.4)

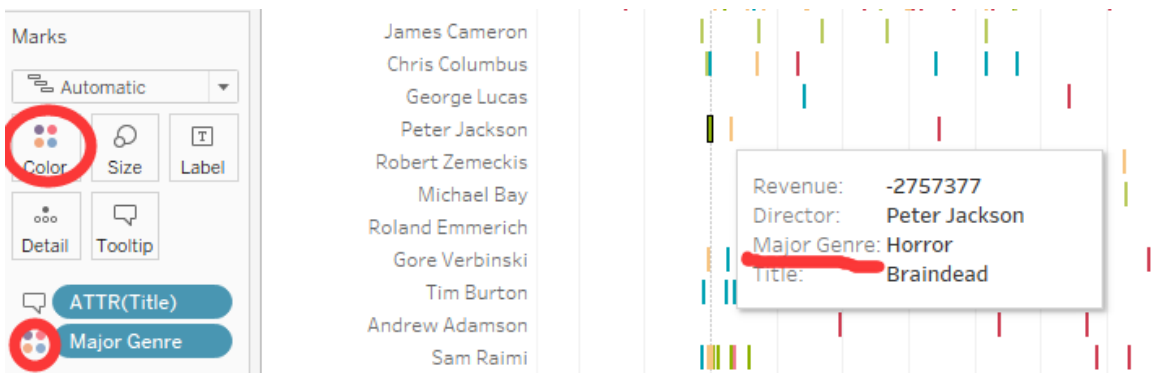


Figure Q8.4

Question 9: Are there directors who have never directed profitable movie? Who are they?
 (Hint 1: you can use filters to complete this task. Hint 2: never directed profitable movie = all movies directed have revenue less than zero)

We want a visualization that either filters these or pulls them “to the top” of the visualization.

Question 10: Can I get more information about a specific genre?

- 1) Go back to the “Sum and Avg of Revenue” worksheet. Drag “Major Genre” to the “Filter” shelf and display only movies from the genre “Action”. (Figure Q10.1)

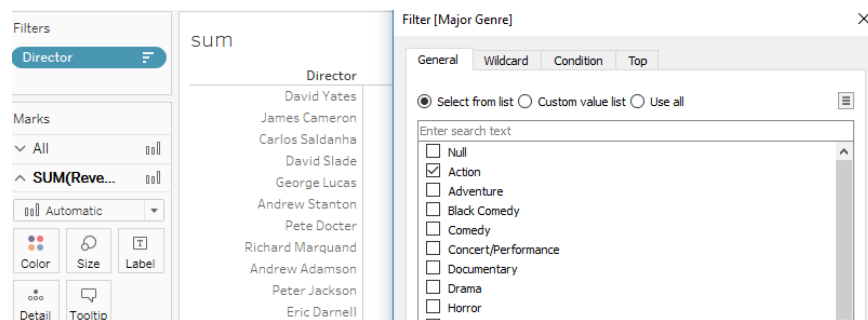


Figure Q10.1

- 2) Find Directors who have the top 10 DVD sales by adjusting the director filter (Figure Q10.2).

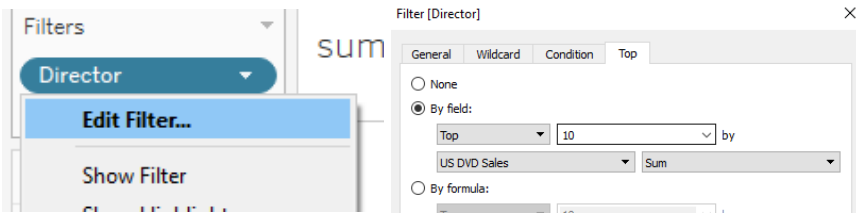

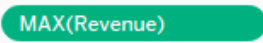


Figure Q10.2

- 3) Find Directors who have the top 10 DVD sales, and have directed action movies

Question 11: Which director has directed the most profitable Comedy movie?

(Hint: use  Columns  and filters/pages shelf)

Question 12: (bonus question): Do movies that contain the word “star” have higher average revenue than movies with “dragon” in the title?

(Hint: create a new group field for both star group and dragon group)

Question 13: When is the best quarter to release a movie?

- 1) Create a new worksheet and plot a visualization of revenue by quarters. (Figure Q13.1)

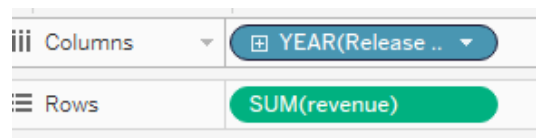


Figure Q13.1

- 2) Click on the plus sign next to the YEAR, or right click on the pill.

PART 3: Putting it all together

Question 14: Can we see directors, movies, and revenue at the same time?

- 1) Go back to the “Avg and Sum” sheet, remove the director filter and sort director names.
- 2) Create a dashboard to host both worksheets that we have created (e.g. “Sum and Avg of Rev” and “Rev as Dimension”) (Figure Q14.1).

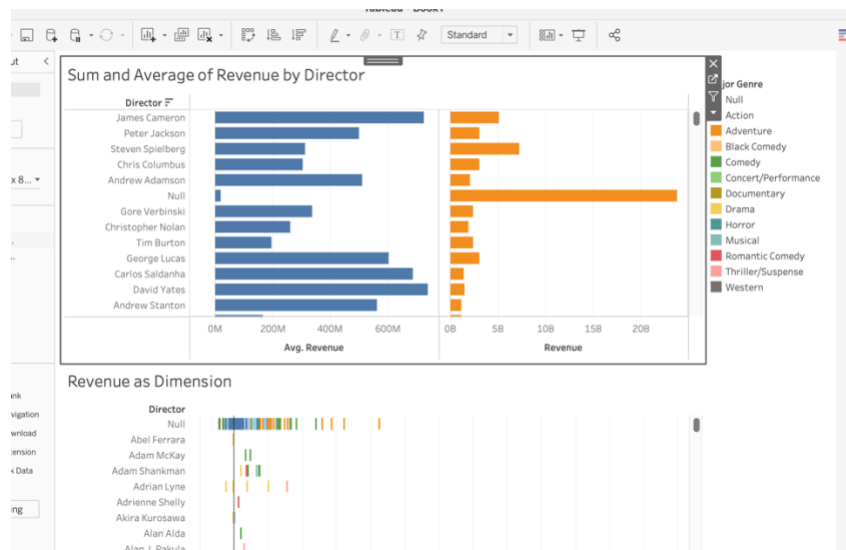


Figure Q14.1

- 3) Add a highlight action: "Dashboard"-> "Actions" -> "Add action"-> "Filter". In the pop-up window, add a "Hover" action for the field of "Director". (Figure Q14.2).

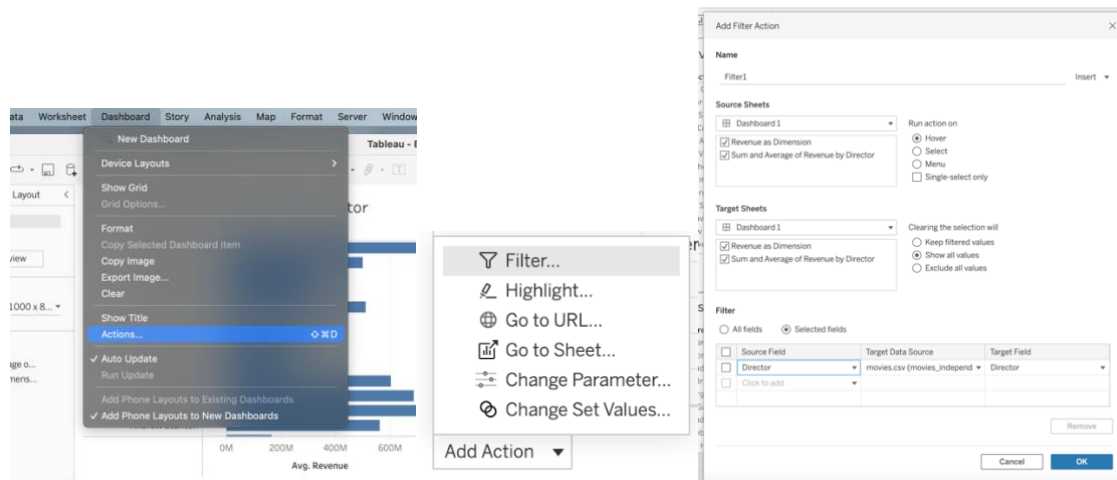


Figure Q14.2

Question 15: Is it better to invest in action movies or comedies? Which director would you invest in?

(This is an open-ended question. Please write down the steps that you used to create the visualizations, and describe how these visualizations help you to make your decision)

Question 16: What's the relationship between IMDB ratings and revenue? Are highly-rated movies more profitable?

(Please write down the steps that you used to create the visualizations, and describe how these visualizations help you to make your decision)

Question 17: Can we get this in a story form?

- 1) Make a story by clicking on the “add story” button (Figure Q17.1)

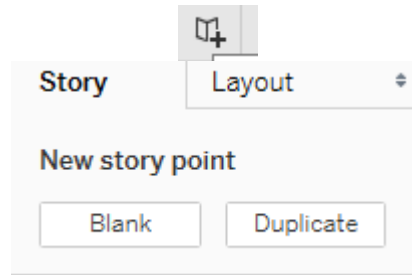


Figure Q17.1

- 2) Annotate the story and add new story points.

Question 18: Who would you invest in?

Create some visualizations to justify your decision. You can use the ones you’ve already created as a start, but we’ll give extra credit for some creativity here.

PART 4: Exporting your visualizations

There are various ways to export your work (including images and PDF). You can also export Interactive Tableau workbooks. This can only be done to Tableau Cloud, Tableau Server and Tableau Public. However, you can embed these visualizations to your own website: see the instruction here: https://onlinehelp.tableau.com/current/api/js_api/en-us/JavaScriptAPI/js_api.htm.

For image export, you can only export worksheets: (Figure 5)

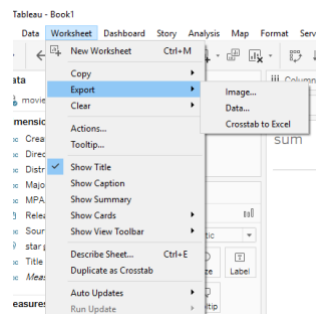


Figure 5: export by clicking worksheet-> export->image

For PDF export (which you can edit on editors such as illustrator), go to file-> print to pdf (Figure 6)

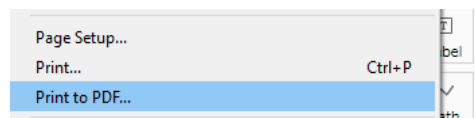


Figure 6

NOTE: Make sure that you check out the tableau public gallery <https://public.tableau.com/en-us/s/gallery>

Independent Practice

“The Dollar-And-Cents Case Against Hollywood’s Exclusion of Women”

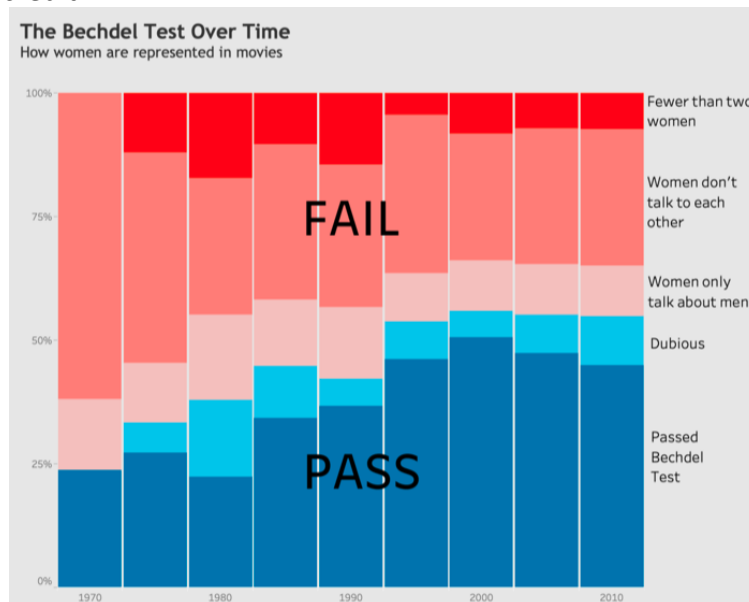
Read this first: <https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/>

(A copy of this article is also available on canvas: The Dollar-And-Cents Case Against Holly...pdf)

Tableau, and tools like it, are designed to help users quickly create visualizations for communication. In this task, you will use Tableau **to create visualizations that communicate and complement the key points** made in the article.

Your task will be to:

- ☐ **Replicate the three** visualizations in the original article using Tableau (*work to get these as close in look and feel as you can... fonts, colors, etc.*) Here’s an example created in Tableau that will earn you full credit. You can get even closer to the original for extra credit.



and

- ☐ **Add three new** visualizations that enhance the article. We’ve expanded the original 538 dataset to offer additional variables to help you here. There should be a good reason you are adding the visualization to the article. You absolutely should create more than three to find interesting/useful visualizations. Hint for a successful choice: if you can find a place in the text to add “see figure” it’s probably a good start.

Important: *When we grade* we will be looking to see that the visualizations you created “fit” the article text (support or complement what the article says), are well labeled, captioned, annotated, styled, etc. They should look like something you’d see in the original article. *If you happen to find something that contradicts a claim (it’s possible!) add that at the end of the article with a brief explanation.*

What to turn in: A PDF of your improved article (3 replicated original visualizations, 3 new ones) and your Tableau workbook with your work.

Helpful resource

The R script used by fivethirtyeight to create their graphs is here:

<https://github.com/fivethirtyeight/data/blob/master/bechdel/analyze-bechdel.R>