DIY Analytics: Beyond Excel

Presentation Workbook

Introduction

The purpose of this workbook is to provide a hands-on demonstrate some of the basic concepts and techniques discussed in the presentation. All that is required to complete these exercises is a computer with a connection to the Internet and a reasonably modern web browser. The data you will use and the software are free. The goal is to make some of the concepts a little less abstract, and show how easy it can be to glean insights from data.

Workbook Scenario

For this exercise you will be analyzing the research patterns of a fictional law firm based on data collected in 2017. This law firm has approximately 1,300 attorneys spread between offices in Tallahassee, Chicago, and Honolulu. The firm has a team of 15 researchers, who are evenly distributed between the three offices.

All research requests are logged in an internal system that automatically assigns a serial number and logs the basic request details (who made the request, the date it came in, etc.) Researchers who work on requests assign a category, note the amount of time spent working on it, and one of the resources used to answer the request (Lexis, Westlaw, etc.)

Note: The data used in this exercise is a work of fiction. It was generated by a computer program. If you are curious about the program that created this data set, it is available at the same web site as the request data. The names used for attorneys and researchers were selected at random from a list of characters from Shakespeare's plays.

Step 1: What Questions Are We Trying to Answer?

An important first step in any analytics project is having some idea of what problem you are trying to solve. For this exercise we will be looking at three specific questions of interest to the research team:

1. Are there any offices that are underserved by the research team?

2. Are any of the firm's resources underused?

3. What are the researchers spending the most time working on?

During the exercise we may be able to answer these questions. We may also find that further data gathering is needed. This is a normal part of analytics: sometimes you will not have all the raw data you need.

Step 2: Download the Request Data

Data for this exercise is in "CSV" format in a file called "requests.csv". It is available on the Internet at

https://github.com/eyadams/AALL-2021-DIY-Analytics-Beyond-Excel

Here is a sample line from request.csv:

```
3364,2017-01-01,Ariel,Honolulu,Real Estate,Justice
Silence,Tallahassee,Company Research,25,Lexis
```

Each line in the file has 10 data elements, or fields:

- Research Request Number ("3364")
- Date request was submitted ("2017-01-01")
- Name of the attorney who made the request ("Ariel")
- City where the attorney is based ("Honolulu")
- The attorney's practice group ("Real Estate")
- The researcher who worked on the request ("Justice Silence")
- The city the researcher is based in ("Tallahassee")
- The category assigned by the researcher ("Company Research")
- The time spent on the request ("25")
- A resource that was used ("Lexis")

The file covers requests for the entire year. It is around 8 megabytes in size.

Step 3: Create a Trial Account with Tableau

This section take you through the steps of creating an account with Tableau.com and signing up for a free trial of their online product. Tableau is available both as a stand-alone application and an online service; the screen shots in this workbook are from the online version. If you prefer to download and install the stand-alone version of Tableau the interface and some minor details will be slightly different, but you will still be able to follow along with the exercises.

Note: there is also a free version of Tableau, called Tableau Public. For the most part it is identical to the trial version of Tableau, with one import difference: any analyses you save will be published to the Tableau Public web site and viewable to all other users of Tableau Public. If you are only working through the exercise in this workbook, that's fine! But if you feel inspired to analyze your own proprietary data, keep this limitation in mind.

1. Navigate to https://www.tableau.com and click on the orange button in the upper right corner labelled "TRY TABLEAU FOR FREE"

2. Under "Tableau Online: Take it to the Cloud" click on "START FREE ONLINE TRIAL".

3. Click on "START YOUR FREE TRIAL"

4. Enter your name, email address, and other information; click on "REQUEST FREE TRIAL" at the bottom. This will generate an email.

5. Click on "Activate My Site" in the email message.

6. You will be prompted to enter your name, choose a password, give the site a name, and pick a site location. For the name, I recommend using the name of your organization, i.e. "WilliamCofLaw" or "ShakespeareLLC"; for the location pick the one closest to you.

7. Read the Tableau Online Subscription Agreement, Data Processing Addendum, and the Terms of Service. Check the box that indicates you have done so and click on "Activate My Site"

Step 4: Create a Workbook and Upload the Data

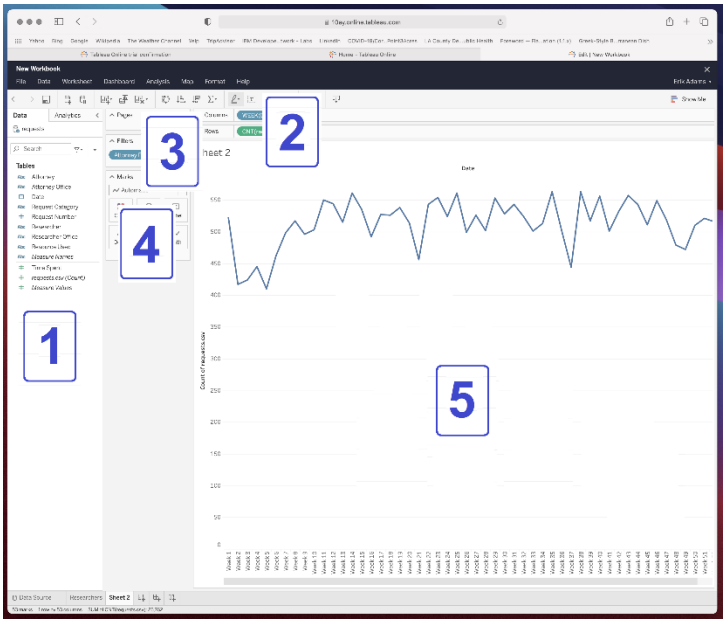
Now that you have an account, you can get started creating a Tableau workbook and set up the data. Tableau makes this process very easy.

1. On the Tableau Online home screen, click on Create → Workbook.

2. On the "Connect to Data" screen, click on the "Files" tab, and then "Upload from computer". In the file selector dialog, find request.csv and upload it.

3. Save your workbook by clicking on File → Save As. Give the workbook a name. Tableau will not save the workbook for you.

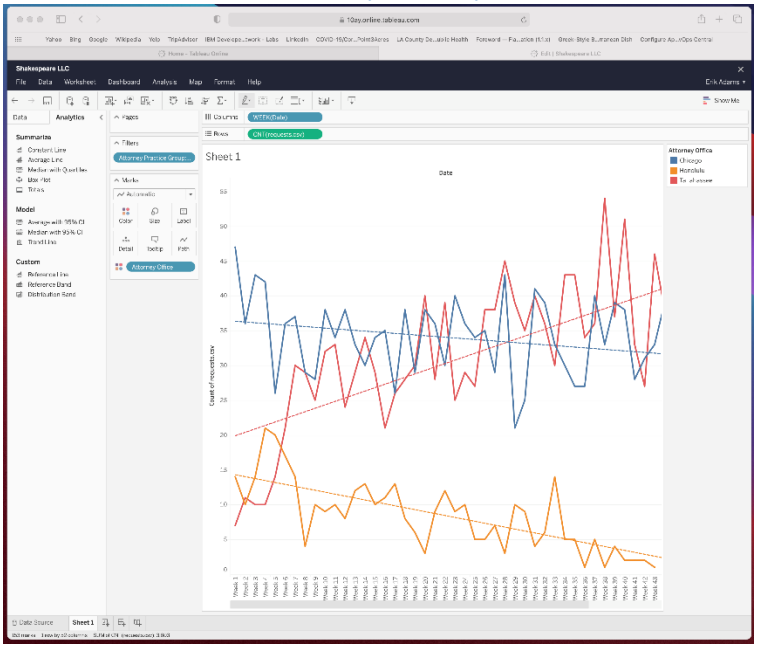
Step 5: Familiarize Yourself with the Tableau Workbook Interface



In Tableau, analysis is done by creating "visualizations" in a "workbook". In addition to the normal menu and tool bar, the Tableau workbook interface has several important parts:

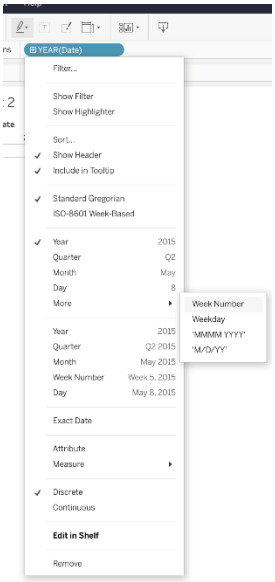
1. The Data and Analytics Pane. The Data tab lists the fields in request.csv, along with some extra values that Tableau calculates automatically. The fields are broken into two sections: "Dimensions" and "Measures". Dimensions, like "Attorney", are listed at the top; Measures, like "Time Spent", are listed at the bottom. Some data fields are calculated from other values – the Measure "requests.csv (Count)" is a count of lines in requests.csv, which means it is also a count of requests. The Analytics tab shows various calculations you can add to your graph, like trend lines and averages.
2. The Columns and Rows shelves. This is where you specify what fields you are analyzing. Fields are displayed as pill-shaped icons that can be moved and manipulated.
3. The Filters Shelf. This is where you focus the data, by adding search criteria to the workbook.
4. The Marks cards. This is a collection of controls used to set the look and feel of the visualization.
5. The View. This is where the visualization is displayed. Users of Tableau often shorten "visualization" to "viz". `

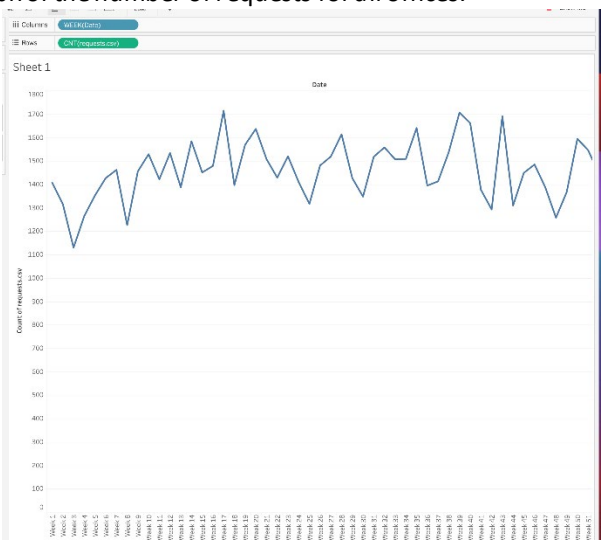
Step 6: Create a Visualization of Requests per Office



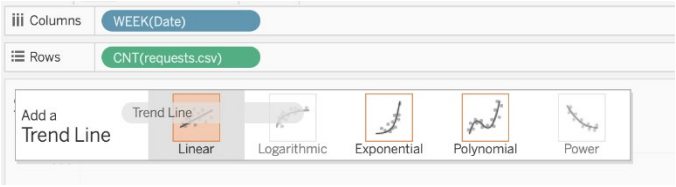
An important first step in analytics is looking for patterns and trends. We will start by creating a simple graph of the number of requests, per week, received by each office in our fictional law firm. When you have completed this section, your visualization will look like the image above.

1. Drag the "Date" dimension to the Columns shelf.
2. Tableau automatically groups dates by year. We want to change this to group by week. To do that, click on the menu arrow at the right of the pill (this can be a little difficult to see), then More → Week Number, as displayed on the right.
3. Confirm that the view now has "Week 1, Week 2, Week 3 …" along the horizontal axis.
4. Drag the "requests.csv (Count)" measure to the Rows shelf.
5. Now you have a graph of the number of requests for all offices:





6. Click on the Analytics tab.
7. Drag "Trend Line" from the pane to the upper right corner of the visualization. A menu will appear with several options, depicted below. Drop the Trend line pill on "Linear"



8. Drag the "Attorney Office" Dimension to the "Color" card. Tableau will automatically assign a color to each office, and then update the visualization so that instead of one graph and one trend line, there will be three. It will also add a key to the right, so that you know which color is for which office.
9. (Optional). Double click on "Sheet 1" and change the name of this visualization to something more memorable, like "Offices Over Time".
10. Click on File → Save to save the visualization. Remember, Tableau does not save your worksheet automatically.

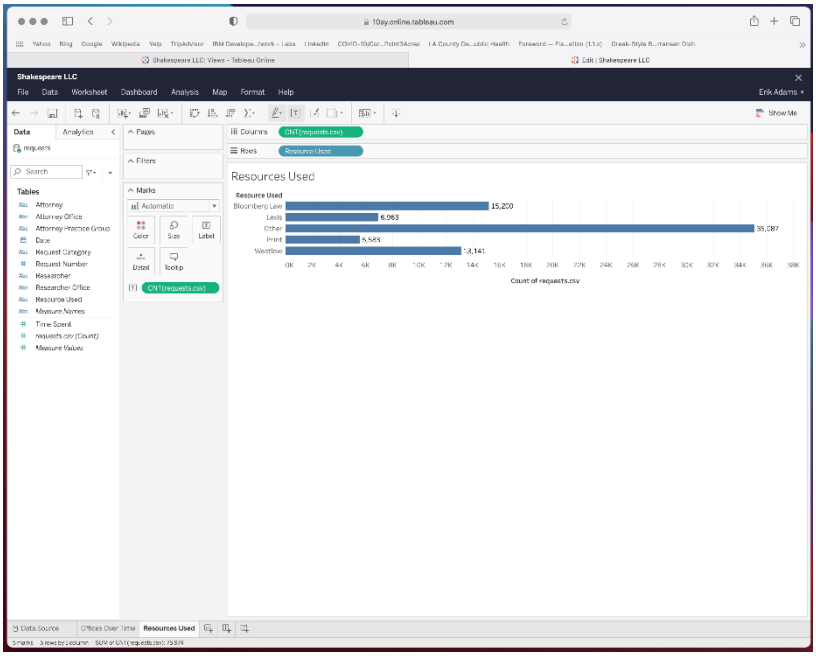
Step 7: Explore the Data

Now that we have a visualization, we can explore the data a little and start looking for insights.

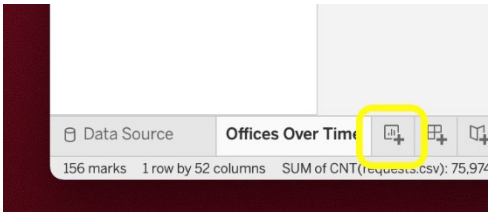
1. Click on "Chicago" in the key at the right of the visualization. The graph and trend lines for the Chicago office will be highlighted, and the other lines dimmed. This makes it a little easier to see that the number of requests originating in the Chicago office is trending down, though not very quickly.
2. Click on "Honolulu" in the key at the right. With the Honolulu trend line highlighted, you can plainly see that the number of requests originating in that office started low at the beginning of the year and was steadily trending down throughout the year.
3. Lastly, click on "Tallahassee". The number of requests from this office started relatively low at the beginning of the year, but noticeably spiked upward in February (between week 4 and week 8) The upward trend continued throughout the year.

At this point, it would be worthwhile to investigate why the number of requests originating in Honolulu and Tallahassee changed so significantly throughout the year. It could be something relatively simple, like the number of attorneys changing in the office, or a shift in the type of law practiced in that office. It could also indicate that the research staff needs to do some outreach to the attorneys in the Honolulu office and make them aware of the range of services provided by the team.

Step 8: Visualize the Use of Resources



For the next question, we will create a new visualization and add new data.



1. At the bottom of the visualization are three buttons with a small plus sign. The one at the left creates a new worksheet, where we will create the new visualization. The image at the right has the correct button highlighted. Click on that button.
2. Drag the "Resource Used" field to the Rows shelf. Five items will be listed: Bloomberg Law, Lexis, Other, Print, and Westlaw.
3. Drag the "requests.csv (Count)" to the Columns shelf.
4. Drag "requests.csv (Count)" to the "Label" card.
5. (Optional) Double click on "Sheet 2" at the bottom and rename the workbook "Resources Used".
6. Save the visualization.

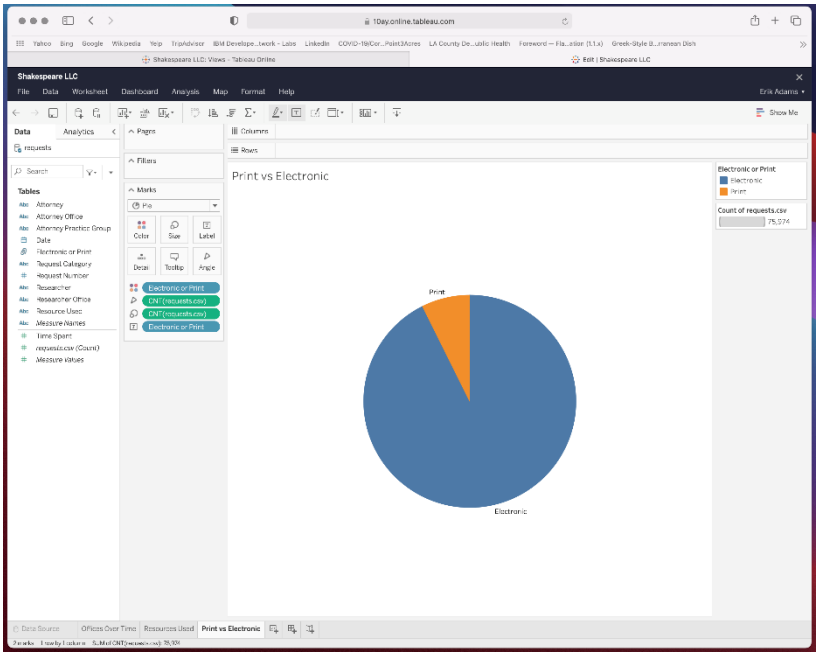
Step 9: Interpret the Results

There are a couple of things worth noting about this visualization. First, it appears that print resources are being used much less frequently than other resources. Most librarians will probably not find this surprising, I'm sure. In the next step we will create a new graph that directly compares print and electronic resources that will make this contrast even more obvious.

The other item worth noting is that researchers selected "Other" more than any other resource type. This should be investigated further. It may be that researchers are not accurately recording the resources they use; it may also mean that there are resources being used that should have their own category for researchers to select.

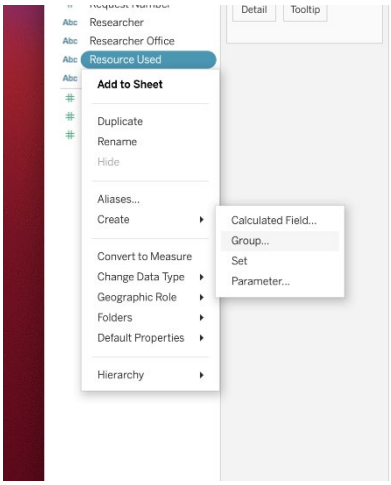
In the next section, we will make a pie chart to emphasize dramatic difference between electronic and print resources.

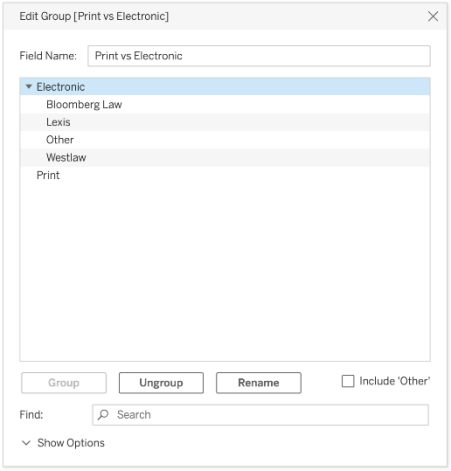
Step 10: Print Versus Electronic



This visualization has a few more steps than the previous ones, but the end result is worth it.

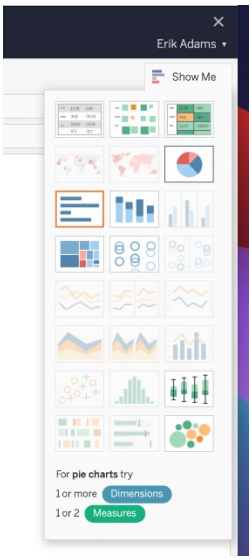
1. Create a new worksheet
2. In the Data pane, click on Resource Used. Then click on the menu arrow → Create → Group. The menu is displayed at the right.
3. This will open the "Edit Group" Dialog, pictured below.
4. In the "Edit Group" dialog, change the field name to "Electronic or Print"
5. Control-click on Bloomberg Law, Lexis, Other, and Westlaw.
6. Click on "Group" at the bottom of the dialog.
7. This will create a new group called "Bloomberg Law, Lexi, Other and 1 more". Click on the group name, and then "Rename".
8. Name the group "Electronic"



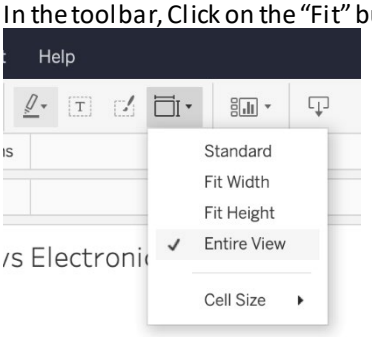


9. Close the Edit Group dialog.
10. Drag the "Electronic or Print" field from the data pane to the columns shelf.
11. Drag the "requests.csv (Count)" field from the data pane to the columns shelf.

12. At the right of the tool bar, click on “Show Me” menu, and on the pie chart



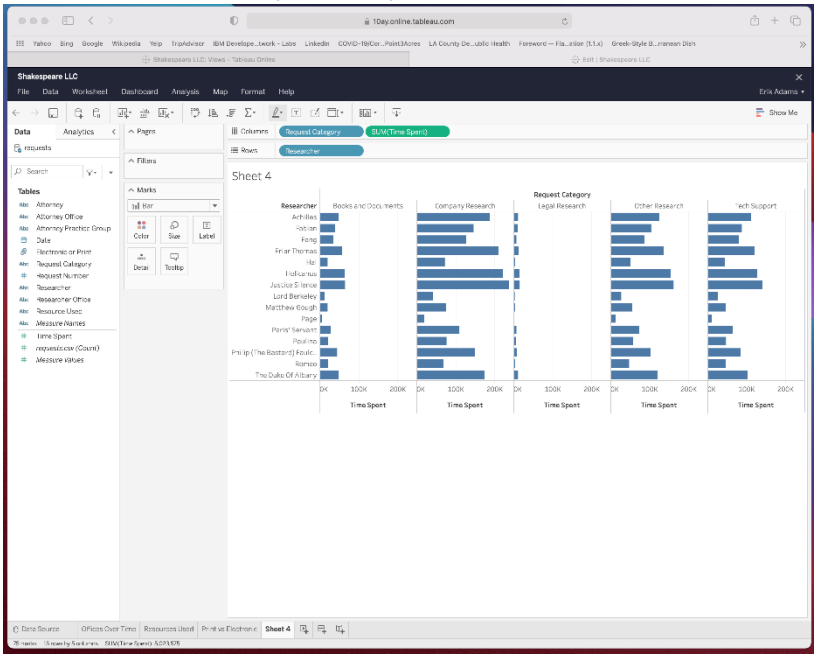
13. In the tool bar, Click on the “Fit” button, and select “Entire View”



14. Drag the “Electronic or Print” field to the Label card.

The pie chart is now complete! A pie chart is a good way to emphasize relative sizes of different data points.

Step 11: Visualize the Time Spent by Researchers



This is the final visualization of this workbook.

1. Create a new workbook.
2. Drag the “Request Category” field from the Data pane to the Columns shelf.
3. Drag the “Time Spent” field from the Data pane to the Columns shelf. Tableau will automatically make this into the sum of time spent.
4. Drag the “Researcher” field from the Data pen to the Rows shelf.
5. (Optional) Rename the worksheet “Researcher Time Spent”

Step 12: Look for Patterns and Trends

The first thing to note about this graph is that the researchers at this firm are spending very little time doing legal research. They are spending a much larger amount of time on company research. This is likely due to patterns in the requests from attorneys.

Another important thing to note is that the researchers appear to be spending almost as much time providing tech support as they are other kinds of research. This would merit further investigation, to see what kinds of technical support are being provided and to see if some of that work can be shifted to another department.

Lastly, it appears that Page is doing much, much less work than everyone else. There may be legitimate reasons for that, but it would be worthwhile to know them and to note it.