



Data Analytics in Accounting: Visualizing Corporate Income Inequality

Seungjae Shin

Mississippi State University, Meridian, sshin@meridian.msstate.edu

Kevin Ennis

Mississippi State University, Meridian, kennis@meridian.msstate.edu

Abstract

Accounting courses increasingly require students to learn data analytics skills and extract insights from business data. The issue of corporate income inequality (the growing gap between the most- and least-profitable companies) has become relevant in discussions of global economic issues. This project has students use data analytic software tools to gain insight into corporate income inequality by extracting and analyzing accounting data. Students learn to clean, merge, and manipulate financial statement data sets to identify corporate income inequality through visualizations. Students' responses to the post-project questionnaire show that the project provided a positive learning experience and increased their academic performance.

Keywords

Accounting information systems, data analytics, corporate income inequality, data visualization, Excel, Tableau

This project, included in the Appendix, guides accounting students as they conduct an exploratory analysis of corporate income inequality, a significant global economic issue. Students first clean, merge, and analyze accounting data sets. Then, they use those data sets to identify the economic issue through data visualizations. The data used include the financial statements of U.S. publicly traded S&P 500 companies. The example we present here includes financial statements from 2001 through 2015. Students use MS Excel and Tableau to identify overall income inequality and to drill down to an industry-level analysis of two sectors, manufacturing and retail trade.

Prior Literature and Motivation for the Project

AIS education aims to help students understand how information technology can help them collect and analyze accounting data to support management decision-making. Thus, using current business software tools helps accounting students prepare for their careers (Lee, Kerler, and Ivancevich, 2018). The AAA-AICPA Pathways Commission (2015) identified accounting students' top three most-needed technologies: electronic spreadsheets, business intelligence and analytics, and database management.

Data Analytics: Professional Standards and Academic Literature

Accounting Accreditation Standard A5, issued by the Association to Advance Collegiate Schools of Business (AACSB International, 2018), states that data analytics and relevant information technology skills should be included in accounting curricula. The CPA Evolution initiative (Tish & Reeb, 2019) emphasizes the importance of data management and analytics as a core area of knowledge expected in newly licensed professional accountants, with three of its five guiding principles mentioning specific types of technology expertise, including data analytics, robotics and artificial intelligence (Coffey, 2019).

Data Analytics: Professional Literature

Tschakert et al. (2016) exhort practicing CPAs to increase their use of data analytics tools in auditing, financial planning, and analysis. Murphy and Tysiac (2015) argue that accountants are using data analytics tools and improved information technologies to replace traditional audit sampling with continuous testing of data populations. Ovaska-Few (2017) predicts that accounting graduates with data analytics skills will be in high demand but notes that teaching resources for those skills are limited, leading to concerns that accounting curricula today do not provide needed data analytics skills. Hart (2018) and Zhang (2019) provide teaching resources for data analytics in accounting. Hart (2018) notes that large accounting firms now offer data analytics resources to educators. Zhang (2019) argues the urgency of integrating data analytics tools, such as ACL, Alteryx, Excel, IDEA, PowerBI, and Tableau, into the accounting curriculum.

Income Inequality

Income inequality is an income distribution gap between two identifiable groups of people or households (Faello et al., 2018). Furman and Orszag (2015), summarized in Ip (2015), examined firm-level income inequality and found that employees of firms with super-normal returns on capital (ROC) shared in those returns, which increased individual income inequality. Specifically, Ip (2015) described return on capital (ROC) trends for non-financial S&P 500 companies from the 1960s to the 2010s and found the gap in ROC between the 90th and 25th percentiles had been stable from the 1960s through the 1980s, but widened during the 1990s and grew substantially in the first decade of the 2000s.

Project Overview

This section provides an overview of the critical elements of the project. We describe the learning objectives, what knowledge and skills students must have to undertake the project (prerequisites), and discuss issues related to acquiring the data needed. We also note the software licenses required to undertake the project.

Learning Objectives

Our review of the professional and academic literature on data analytics in accounting and the critical global economic issue of corporate income inequality leads us to set five learning objectives for this project. Upon completion of the project, students should be able to do the following:

1. Process (explore/clean/extract) an existing accounting data set
2. Use appropriate software tools, such as MS Excel and Tableau, to transform accounting data into useful information
3. Perform drill-down analysis to find more detailed information
4. Present outputs using visualization software
5. Describe the advantages and disadvantages of using various analytics software, such as MS Excel and Tableau

Prerequisites

This case requires that students know how to read financial statements and perform fundamental ratio analysis. Most business students will learn these skills in an introductory financial accounting course. Accounting students will often have this knowledge reinforced in subsequent intermediate financial accounting and auditing courses.

Students will also need basic knowledge of MS Excel, how its basic and intermediate functions (such as VLOOKUP, COUNTIF, and LEFT) work, and how to filter data in the software. Business students often learn these skills in an introductory, lower-division, electronic spreadsheets course.

Basic Tableau skills (such as converting data types, filtering data, creating hierarchy measures, creating calculated fields, and creating various charts) are also prerequisites for undertaking the project. Although many business schools offer training in these skills, most do not. We used the Tableau (2018) training site, which provides free tutorial videos for the skills needed to complete the project.

Data and Software License Sources

We obtained annual financial statements for U.S. publicly traded companies from Compustat, which we accessed through the Wharton Research Data Services (WRDS) interface and downloaded to MS Excel. For adopters at schools without WRDS access, an alternative free source of the most recent five years of historical financial statements is Morningstar.com. Adopters can work with their institution's business librarian to identify other sources of recent financial statements.

Microsoft Excel is widely available to university students at low or no cost through school licensing programs or the Microsoft Education (2022) Web site. Tableau (2022) provides free, limited-term academic student licenses. We created this project using the MS Excel 2016 and Tableau 2019.2 versions.

Efficacy

We have used the project in undergraduate accounting information systems courses and graduate management information systems and supply chain management courses over two years. These courses have a data analytics component and include using MS Excel and Tableau software. To learn more about the project's efficacy, we surveyed participating students in compliance with the rules of the relevant institutional review board. We first asked students to estimate their pre-project MS Excel and Tableau skill levels and rate themselves on a five-point scale (1, no experience; 2, beginner; 3, intermediate; 4, advanced; 5, expert). We report their responses in Table 1.

Table 1

Pre-project Students' Self-assessed Software Skill Levels^a

Software	Undergraduate Students (n = 20)		Graduate Students (n = 57)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
MS Excel	2.65	0.xx	3.12	0.73
Tableau	1.05	0.xx	1.56	0.85

^aFive-point scale: 1, no experience; 2, beginner; 3, intermediate; 4, advanced; 5, expert

Students reported having more expertise in MS Excel than Tableau, which is what we expected. Most of the students had never used Tableau before.

We also asked participating students to rate four statements about their experience using the software and their reaction to doing the project. We asked them to use a five-point scale (1, not at all; 2, slightly; 3, somewhat; 4, very; 5, extremely) to indicate their level of agreement with the statements, as summarized in Table 2.

Table 2
Students' Agreement with Statements about their Experience with the Project^a

Statements	Undergraduate		Graduate	
	Students (n = 20)		Students (n = 57)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Using Tableau helped me understand data analytics better	4.15	1.04	3.98	0.86
The project was interesting	3.60	0.82	3.77	0.66
The project was difficult	3.00	0.97	2.58	0.89
The project was a positive learning experience	4.05	1.00	4.11	0.70

^aFive-point scale: 1, not at all; 2, slightly; 3, somewhat; 4, very; 5, extremely

The overall perceptions about this case project (that it was interesting, helpful, and a positive learning experience) were encouraging. Responses from graduate students were higher than those from undergraduate students, except for their perceptions of the degree of difficulty. Because we created this case to demonstrate the use of software tools in data analytics rather than to compare one tool to another, we did not compare student performance with a pre-test and post-test.

Course grades (of which the project grade was ten percent) for the 77 students who completed the survey included 25 As, 46 Bs, and 6 Cs. Student assessment of a “positive learning experience” has a statistically significant relationship (correlation coefficient of 0.265, $p = 0.02$) with students’ final grades when we compare students earning As with those earning other grades (Bs or Cs).

Conclusion

Lawson (2019) finds that managers increasingly rely on accounting teams to provide strategic data analyses; thus, a shortage of accountants with those skills is an obstacle to getting good strategy support. Therefore, data analytics education in the accounting curriculum is critical. This project helps students use analytics software tools (MS Excel and Tableau) to extract, clean, and analyze useful financial statement data information.

References

- AACSB International. (2018). Eligibility procedures and accreditation standards for accounting accreditation. <https://www.aacsb.edu/accreditation/standards/accounting>
- AAA-AICPA Pathways Commission (2015). In pursuit of accounting’s curricula of the future. American Accounting Association (AAA)-American Institute of Certified Public Accountants. <https://aaahq.org/portals/0/images/education/pathways/15-9-61866.pdf>
- Coffey, S. (2019, July 24). 5 principles for evolving CPA licensure. *AICPA Insights*. <https://blog.aicpa.org/2019/07/5-principles-for-evolving-cpa-licensure.html#sthash.AD2KLA9V.dpbs>
- Faello, J. Qu, Y., Jain, A., Pomare, C. & Shin, S. (2018). Income inequality: Does corporate income inequality parallel individual income inequality? *Journal of Accounting and Finance*, 18(9), 39–78. <https://doi.org/10.33423/jaf.v18i9.123>
- Frick, W. (2016, May 11). Corporate inequality is the defining fact of business today. *Harvard Business Review*. <https://hbr.org/2016/05/corporate-inequality-is-the-defining-fact-of-business-today>
- Furman, J. and Orszag, P. (2015, October 16). A firm-level perspective on the role of rents in the rise in inequality. Columbia University. <http://gabriel-zucman.eu/files/teaching/FurmanOrszag15.pdf>
- Furman, J. and Orszag, P. (2018, June). Slower productivity and higher inequality: Are they related? Peterson Institute for International Economics. <https://www.piie.com/system/files/documents/wp18-4.pdf>
- Hart, L. (2018, June 12). Educator resources from the big four. *Journal of Accountancy Extra Credit Newsletter*. <https://www.journalofaccountancy.com/newsletters/extra-credit/accounting-education-resources-from-big-four.html>
- Ip, G. (2015, November 4). Behind rising inequality: More unequal companies. *The Wall Street Journal*. <https://www.wsj.com/articles/behind-rising-inequality-more-unequal-companies-1446665769>
- Lawson, R. (2019, April). Management accounting competencies: Fit for purpose in a digital age? Institute of Management Accountants. <https://www.imanet.org/-/media/4cab087f40b54bac878cae99c892d9a4.ashx>
- Lee, L, Kerler, W. & Ivancevich, D. (2018). Beyond Excel: Software tools and the accounting curriculum, *AIS Educator Journal*, 13(1), 44–61.
- Microsoft. (2022) Office 365 education. Microsoft Education. <https://www.microsoft.com/en-us/education/products/office>
- Murphy, M. & Tysiac, K. (2015, April 13). Data analytics helps auditors gain deep insight. *Journal of Accountancy*. <https://www.journalofaccountancy.com/issues/2015/apr/data-analytics-for-auditors.html>
- Orszag, P. (2015, October 16). People are not unequal; companies are. The Brookings Institution. <https://www.brookings.edu/opinions/people-are-not-unequal-companies-are/>
- Ovaska-Few, S. (2017, December 12). Prepare accounting students for working with data analytics. *Journal of Accountancy*. <https://www.journalofaccountancy.com/newsletters/extra-credit/prepare-students-data-analytics.html>

- Tableau (2018). Tableau free training videos. <https://www.tableau.com/learn/training/>
- Tableau (2022). Tableau for students. <https://www.tableau.com/academic/students>
- Tish, L. & Reeb, W. (2019, November). CPA Evolution. National Association of State Boards of Accountancy. <https://nasba.org/app/uploads/2019/11/CPA-Evolution.pdf>
- Tschakert, N., Kokina J., Kozlowski, S. & Vasarhelyi, M. (2016, August 1). The next frontier in data analytics, *Journal of Accountancy*. <https://www.journalofaccountancy.com/issues/2016/aug/data-analytics-skills.html>
- The Pathway Commission. (2015, November). In Pursuit of Accounting's Curricula of the Future. The American Accounting Association. http://commons.aaahq.org/files/3968045e18/Pathways_Report_4_For_Print.pdf
- Zhang, J. (2019, June 11). Resources for teaching data analytics in accounting. *Journal of Accountancy*. <https://www.journalofaccountancy.com/newsletters/extra-credit/teaching-data-analytics-in-accounting.html>

Appendix

Data Analytics Project: Corporate Income Inequality

Project Setting

A *Wall Street Journal* (WSJ) article related to global economic issues, “Behind Rising Inequality: More Unequal Companies,” was published on November 4, 2015. The author compared trends in return on capital (ROC) of non-financial companies in the S&P 500 since the 1960s to explore possible corporate income inequality, defined as the growing gap between the most- and least-profitable companies. The results indicate a gap, between the ROC of companies in the top 90% and the ROC of the rest of the S&P 500 companies. The gap started to surge in the 1990s and continues to grow substantially into the 21st century.

Further analysis of this data by Furman and Orszag (2018) suggests that productivity growth has been uneven among economic sectors. For example, the ROC of the top 10% of publicly traded non-financial firms was 20% in the 1980s and 100% or more in the 2010s, indicating that the ROC of the top 10% companies has risen roughly five times over the past three decades. Furman and Orszag (2018) identified two industry sectors, information technology and health care, as contributing the most to income inequality.

Data Description

In this project, you will analyze the annual financial statements for U.S. publicly traded companies from 2001 through 2015. As the WSJ article suggested, the data set excludes the financial industry, using Standard Industrial Classification (SIC) codes between 6000 and 6799, because the ROC data in the financial industry is considerably scarce. There are two worksheets in the student data Excel file: ‘Data’ and ‘S&P500’. The ‘Data’ worksheet has financial statement data from 2001 through 2015 and the ‘S&P500’ worksheet has ticker symbols (column A) and company names (column B), which belong to S&P 500 index. The S&P 500 index is a stock market index, where the 505 common stocks are issued by the 500 largest capitalized companies listed in the U.S. stock markets, such as the NYSE and NASDAQ. The number of records in the ‘Data’ worksheet is 108,156, including header information, and that of ‘S&P500’ is 505, without header information. Table 1 shows 12 data items collected from the annual financial statements.

Table 1

Data items

Data Item	Name	Unit	Type
AT	Total Assets	Millions	US \$
CEQ	Common Stock Equity	Millions	US \$
CN	Company Name		Text
DT	Total Debt	Millions	US \$
DVT	Total Dividend	Millions	US \$
FY	Fiscal Year		Number (yyyy)
GDWL	Goodwill	Millions	US \$
NAICS	North America Industry Classification Code		Number (6 digits)
NI	Net Income	Millions	US \$
REVT	Total Revenue	Millions	US \$
SEC	Security Exchange Code		Number (2 digits)
TS	Ticker Symbol		Text

Project Instructions

Part 1. Understanding the issues about income inequality

Please read the article, Behind rising inequality: More unequal companies, published on November 4, 2015 (<https://www.wsj.com/articles/behind-rising-inequality-more-unequal-companies-1446665769>). If that article is not available, the following are alternatives:

Orszag, P. (2015). People are not unequal; companies are. <https://www.brookings.edu/opinions/people-are-not-unequal-companies-are/>

Frick, W. (2016). Corporate inequality is the defining fact of business today. <https://hbr.org/2016/05/corporate-inequality-is-the-defining-fact-of-business-today>

Part 2. Processing the data set with Excel

Your instructor will provide the source data set (Income-Inequality-Student-Data.xlsx), which includes two worksheets of ‘data’ and ‘S&P500’. In this part, students learn the data processing technique from raw data to organized data, such as data verification, data cleaning, merging & sorting, to readying for a further data analysis. Figure 1 & 2 show the first several lines of data in each worksheet. In addition to the two worksheets, there are several blank worksheets with tab names of ‘Count’, ‘Data-S&P500’, ‘S&P500-Only’, ‘S&P500-Not-Null’, ‘S&P-500-Sector’ and ‘Sector’ (Figure 3) At the end of this project, students need to add three additional worksheets with tab names of ‘ROC-Export’, ‘ROC-Sector-Export’ and ‘Memo’. Table 2 explains each Excel worksheet.

Figure 1

Financial data (2001~2015)

	A	B	C	D	E	F	G	H	I	J	K	L
1	FY	TS	CN	AT	CEQ	DT	DVT	GDWL	NI	REVT	SEC	NAICS
2	2015	AIR	AAR CORP	1442.1	865.8	148.1	10.4	117.3	47.7	1662.6	11	423860
3	2015	AVX	AVX CORP	2409.819	2177.106	0	70.499	213.051	101.535	1195.529	11	334416
4	2015	PNW	PINNACLE	15028.26	4583.917		270.519	0	437.257	3495.443	11	2211
5	2015	AAN	AARON'S	2658.875	1366.618	610.45	6.822	539.475	135.709	3179.756	11	532299
6	2015	ABT	ABBOTT L	41247	21211	5874	1464	9638	4423	20405	11	325412
7	2015	AEM	AGNICO E	6683.18	4141.02	1151.774	68.762	696.809	24.583	1985.432	11	212221
8	2015	APD	AIR PROD	17438.1	7249	4384.7	687.9	1131.3	1277.9	9894.9	11	325120

Figure 2

S&P500 company list

	A	B
1	A	Agilent Technologies
2	AAL	American Airlines Group
3	AAP	Advance Auto Parts
4	AAPL	Apple
5	ABBV	AbbVie
6	ABC	AmerisourceBergen
7	ABMD	Abiomed
8	ABT	Abbott Laboratories

Figure 3

Names of student Excel worksheets

Data	Count	S&P500	Data-S&P500	S&P500-Only	S&P500-Not-Null	S&P500-Sector	Sector
------	-------	--------	-------------	-------------	-----------------	---------------	--------

Table 2

Description of Excel Worksheets

Name of Worksheet	Data Description in Worksheet
[1] Data	Source data for financial statement items
[2] Count	Summarized table for count of records by year
[3] S&P500	Company list of S&P500 Index
[4] Data-S&P500	Source data with ‘S&P500’ field
[5] S&P500-Only	‘Data-S&P500’ data without Non-S&P500 companies
[6] S&P500-Not-Null	‘S&P500-Only’ data without blank cells
[7] S&P500-Sector	‘S&P500-Not-Null’ with ‘Sector’ field
[8] Sector	Summarized table for count of records by sector

Step 1: Summarize the count of records by year

Q1 How many records in each year from 2001 through 2015 respectively are in the ‘Data’ worksheet? Summarize the count of records by year on ‘Count’ worksheet.

Once you open the data file ('income-inequality-student-data.xlsx'), make sure the data set is sorted by FY. Count the number of data records in the 'Data' worksheet using the 'countif()' function. Save the output in 'Count' worksheet (Figure 4).

Figure 4

Count worksheet

Q2 What is the trend of number of data from 2001 through 2015? Is it a pattern of increasing, decreasing, or constant? Select the data you counted in the 'Count' worksheet and insert a bar chart with a trend line.

Step 2: Delete records that do not correspond to S&P500 companies, are duplicates or are null records.

Copy all data in the worksheet 'Data' and paste them in 'Data-S&P500' worksheet. Add a new column, 'S&P500' (column M), using the 'vlookup()' function. The 'vlookup()' function compares the TS (ticker symbol) of 'Data' worksheet (Figure 1) with ticker symbol column (no column header) in the 'S&P500' worksheet (Figure 2). If a value of TS in the 'Data' worksheet matches with of the value in the column of ticker symbol (column A) in the 'Data-S&P500' worksheet, put a value of TS in the column N. And also, using the 'iferror()' function together with 'vlookup()' function, if no matching of ticker symbol, type '0' in the cell (Figure 5).

Figure 5

S&P500 column in 'Data-S&P500' worksheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FY	TS	CN	AT	CEQ	DT	DVT	GWDL	NI	REVT	SEC	NAICS	S&P500
2	2015	AIR	AAR CORP	1442.1	865.8	148.1	10.4	117.3	47.7	1662.6	11	423860	0
3	2015	AVX	AVX CORP	2409.819	2177.106	0	70.499	213.051	101.535	1195.529	11	334416	0
4	2015	PNW	PINNACLE	15028.26	4583.917		270.519	0	437.257	3495.443	11	2211	PNW
5	2015	AAN	AARON'S	2658.875	1366.618	610.45	6.822	539.475	135.709	3179.756	11	532299	0
6	2015	ABT	ABBOTT	41247	21211	5874	1464	9638	4423	20405	11	325412	ABT

Q3 How many '0' value / non-zero value are in the column M (S&P500) of the 'Data-S&P500' worksheet. Use a countif() function to find the number of '0' value and non-zero values in the column M.

Select all data using Excel short cut "Ctrl+A" in the worksheet 'Data-S&P500' and make a table using 'Format as Table' button in the 'Home' menu ribbon / 'Styles' group (Figure 6) Click the drop-down arrow in column M (S&P500) and unselect "0" (Figure 7) Copy and paste the filtered records into a new worksheet with a name of 'S&P500-Only'. Make sure that you select a paste option of 'Values' in applying 'paste' command (Figure 8)

Figure 6

Format as table button

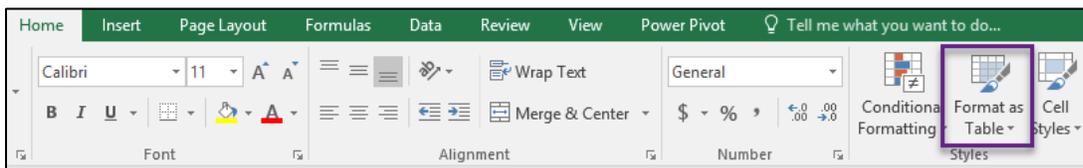
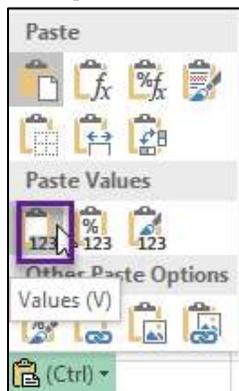


Figure 7
S&P500 column

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FY	TS	CN	AT	CEQ	DT	DVT	GDW	NI	REVT	SEC	NAIC	S&P500
2	2015	AIR	AAR CORP	1442.1	865.8	148.1	10.4	117.3	47.1				
3	2015	AVX	AVX CORP	2409.819	2177.106	0	70.499	213.051	101.53				
4	2015	PNW	PINNACLE	15028.26	4583.917		270.519	0	437.25				
5	2015	AAN	AARON'S	2658.875	1366.618	610.45	6.822	539.475	135.70				
6	2015	ABT	ABBOTT L	41247	21211	5874	1464	9638	442				
7	2015	AEM	AGNICO E	6683.18	4141.02	1151.774	68.762	696.809	24.58				
8	2015	APD	AIR PROD	17438.1	7249	4384.7	687.9	1131.3	1277.1				
9	2015	ALK	ALASKA A	6533	2411	686	102	0	84				
10	2015	MATX	MATSON I	1669.8	450.6	429.9	30.8	241.6	10				
11	2015	HON	HONEYWE	49316	18283	6131	1686	15895	476				
12	2015	ARNC	HONCONIC	36528	11988	9065	223	5401	-32				
13	2015	HES	HESS COR	34195	19386	6630	287	375	-305				
14	2015	ABM	ABM INDL	2149.8	1007.5	158	36	867.5	76				
15	2015	AEP	AMERICAN	61683.1	17891.7		1059	52.5	2047				
16	2015	AVD	AMERICAN	443.539	268.395	68.376	0.572	0	6.59				
17	2015	AME	AMETEK I	6664.53	3254.626	1628.02	85.988	2706.633	590.85				
18	2015	AP	AMPSCO-P	506.156	211.423	13.311	7.58	0	1.37				
19	2015	AXR	AMREP CC	120.628	82.844			0	0				
20	2015	AXR	AMREP CC	120.628	82.844	12.939		0	0				
21	2015	NBR	NABORS I	8527.84	4382.71	2661.708	69.262	166.658	272.67				

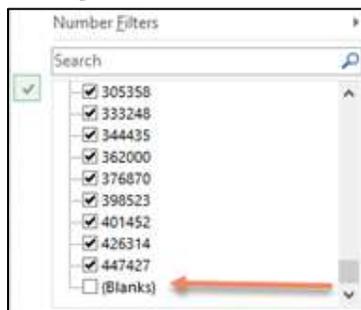
Figure 8
Paste option: 'Values'



Q4 How many data records are in the worksheet 'S&P500-Only' with the header line?

Q5 Can you find empty cells? Which data field has empty cells? Click drop-down arrow of each column header and unselect 'blanks' option (Figure 9).

Figure 9
Filtering blank cells



Copy and paste remaining data into a new worksheet named 'S&P500-Not-Null' (Figure 10) Make sure to select 'Values' option in applying 'paste' command (Figure 8).

Figure 10
'S&P500 Not Null' worksheet

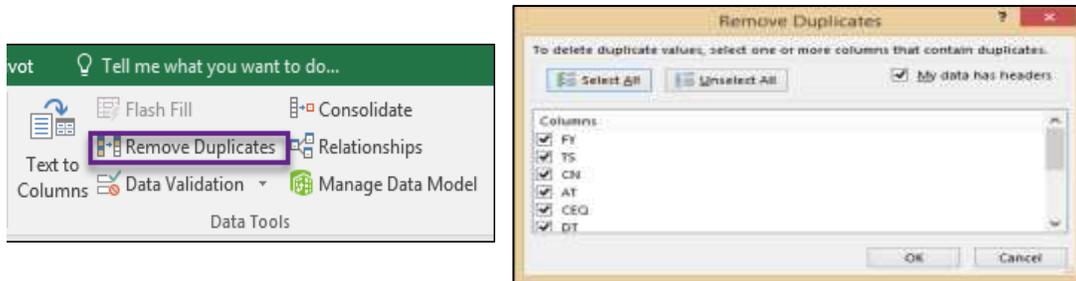
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FY	TS	CN	AT	CEQ	DT	DVT	GDWL	NI	REVT	SEC	NAICS	Sector
2	2015	MON	MONSANTO CO	21920	6990	8737	952	4061	2314	15001	11	111 11	
3	2014	MON	MONSANTO CO	21981	7875	7540	916	4319	2740	15855	11	111 11	
4	2013	MON	MONSANTO CO	20664	12559	2072	831	3520	2482	14861	11	111 11	
5	2012	MON	MONSANTO CO	20224	11833	2042	682	3435	2045	13504	11	111 11	
6	2011	MON	MONSANTO CO	19844	11545	2169	611	3365	1607	11822	11	111 11	
7	2010	MON	MONSANTO CO	17852	10069	2055	583	3204	1096	10483	11	111 11	
8	2009	MON	MONSANTO CO	17877	10056	1755	565	3218	2109	11724	11	111 11	
9	2008	MON	MONSANTO CO	17991	9374	1805	456	3132	2024	11365	11	111 11	
10	2007	MON	MONSANTO CO	12983	7503	1386	299	2625	993	8563	11	111 11	

Q6 How many data records in the worksheet 'S&P500-Not-Null' including header?

Q7 Can you find any duplicate data records in 'S&P500-Not-Null' worksheet?

Select the following menu path to find duplicate data: 'Data' menu ribbon > 'Data Tools' group > 'Remove Duplicate' button (Figure 11).

Figure 11
'Remove Duplicates' button and pop-up



Step 3: Classify record into industry sector using NAICS codes.

Select and copy column A through column L of 'S&P500-Not-Null' worksheet and paste it to 'S&P500-Sector' worksheet. In the column L of 'S&P500-Sector' worksheet, there are NAICS (North America Industry Classification System) codes. These codes are composed of 6 digits. The first two digits represent an industry sector.

Q8 Are there any code whose number is less than 6 digits in the column L (NAICS) of 'S&P500-Sector' worksheet? The NAICS column is the number data type and the minimum number with 6 digit integer number is 100,000. Count the number of NACIS codes using 'countif()' function. Even if the NAICS code should be a 6-digit number, some of NAICS code have less than 6 digits. Add a new column 'Sector' (column M) using a function of left() with the first two digits of the NAICS codes (Figure 12).

Figure 12
Sector column

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FY	TS	CN	AT	CEQ	DT	DVT	GDWL	NI	REVT	SEC	NAICS	Sector
2	2015	MON	MONSANTO CO	21920	6990	8737	952	4061	2314	15001	11	111 11	
3	2014	MON	MONSANTO CO	21981	7875	7540	916	4319	2740	15855	11	111 11	

Copy 'Sector' column (M) of 'S&P500-Sector' worksheet and paste it to Cell A1 in the 'Sector' worksheet with paste option of 'Values' (Figure 8) Click 'Data' ribbon menu and sort it with 'Sort' button in 'Sort & Filter' group and remove duplicate sector codes with 'Remove Duplicates' button in 'Data Tools' group (Figure 13) Add column headers, 'Sector Name' (Cell B1) and 'No. of Records' (Cell C1) in the first row of 'Sector' worksheet. Fill the sector name column (Cell B2 ~ Cell B20) referring to Table 3 (Figure 14).

Figure 13
Sort and remove duplicates

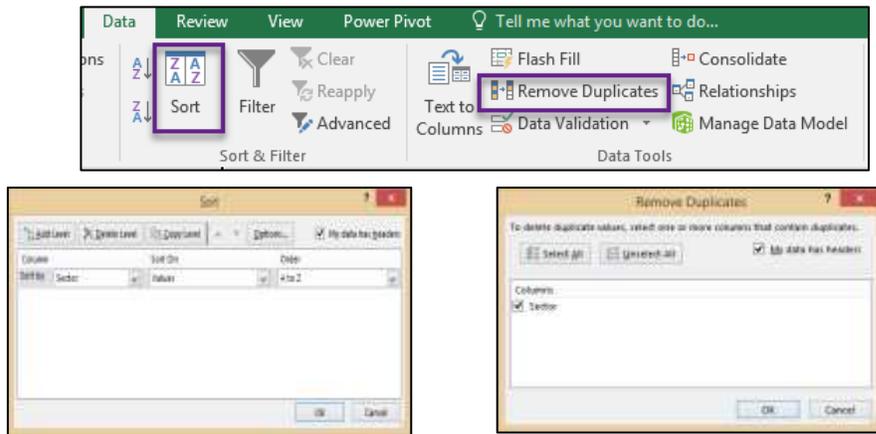


Table 3
NAICS Code and Industry Sector^a

Code	Industry Sector
11	Agriculture, Forestry, Fishing and Hunting
21	Mining
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
92	Public Administration

^aSource: <https://www.naics.com/business-lists/counts-by-naics-code/?#countsByNAICS>

Figure 14
Sector worksheet

	A	B	C	D	E	F
1	Sector Code	Sector Name	No. of Records			
2	11	Agriculture, Forestry, Fishing and Hunting				
3	21	Mining				
4	22	Utilities				
5	23	construction				
6	31	Manufacturing				
7	32					
8	33					
9	42	Wholesale Trade				
10	44	Retail Trade				
11	45					
12	48	Tranportation and Warehousing				
13	49					
14	51	Information				
15	52	Finance and Insurance				
16	53	Real Estate Rental and Leasing				
17	54	Professional, Scientific, and Technical Services				
18	56	Administrative and Support and Waste Management and Remediation Services				
19	62	Health Care and Social Assistance				
20	72	Accomodation and Food Services				
21	99					

- Q9** How many number of data in each sector in the worksheet 'S&P500-Sector'? Put your numbers in the Column C (No. of Record) of the 'Sector' worksheet using a 'countif()' function.
- Q10** Can you find any finance sector record in 'Sector' worksheet? The sector code 52, finance and insurance, is not supposed to be in 'S&P500-Sector' worksheet because the finance industry data should be excluded in the original data set. If 'Yes', what is the ticker symbol for that company?
- Q11** Sort the records in 'S&P500-Sector' worksheet by 'Sector' column (M) with ascending order. Delete the records with a sector code 52 in the 'S&P500-Sector' worksheet. How many records remain in the 'S&P500-Sector' worksheet including column header after deletion?
- Q12** Do you have any record in the sector number 99? The sector code 99 is not in the NAICS code list. (Table 3) If 'Yes', what are ticker symbols for those companies? Even if the code 99 is out of range according to the Table 3, we keep it for further analysis.

Now you have a clean, organized data set ('S&P500-Sector') to upload to Tableau for further analysis. Make sure that there is no data after the last record, row number 4862 (Figure 15).

Figure 15
The Last Part of 'S&P500-Sector' Worksheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FY	TS	CN	AT	CEQ	DT	DVT	GDWL	NI	REVT	SEC	NAICS	Sector
4857	2003	BRK.B	BERKSHIRI	180559	77596	9119	0	22948	8151	63859	11	999977	99
4858	2003	GE	GENERAL	647483	79180	208281	7759	47487	15002	133585	11	999977	99
4859	2002	BRK.B	BERKSHIRI	169544	64037	9288	0	22298	4286	42353	11	999977	99
4860	2002	GE	GENERAL	575244	63706	176306	7266	39138	14118	130685	11	999977	99
4861	2001	BRK.B	BERKSHIRI	162752	57950	12504	0	21407	795	39903	11	999977	99
4862	2001	GE	GENERAL	495023	54824	110838	6555	28287	13684	125679	11	999977	99

Part 3. Visualization with Tableau

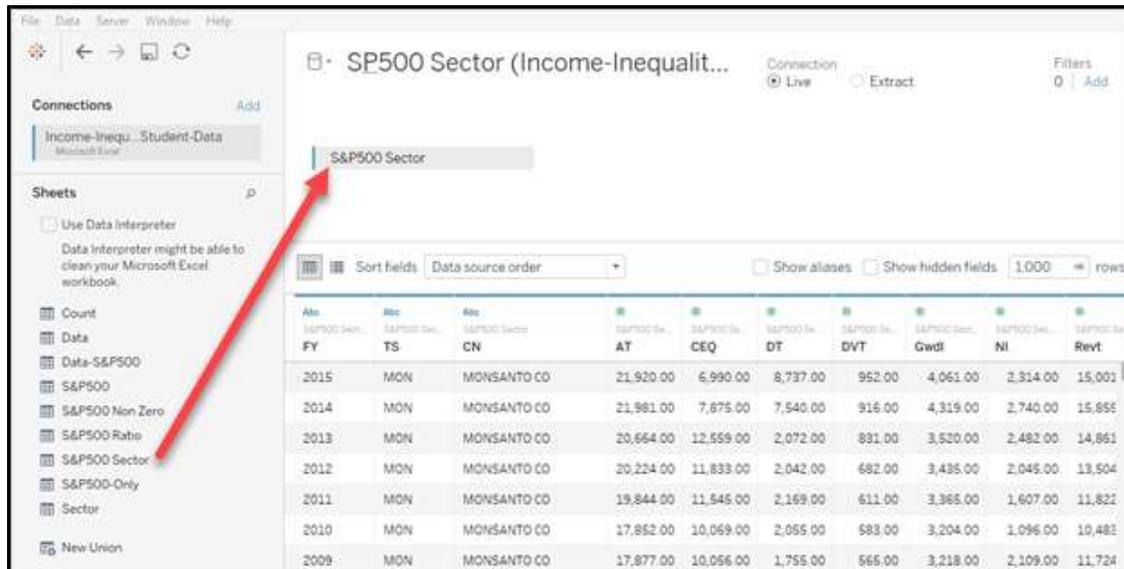
In Part 3, you will learn how to analyze the data to get useful information and how to present output as a form of chart.

Step 1: Upload Excel data to Tableau

Click “Microsoft Excel” under the connect tab and import the ‘Income-Inequality-Student-Data.xlsx’ Excel file to Tableau. In the data source page, drag ‘S&P500 Sector’ worksheet from the left pane to canvas which is located at the upper part of data source page (Figure 16).

Figure 16

Uploading Excel file to Tableau

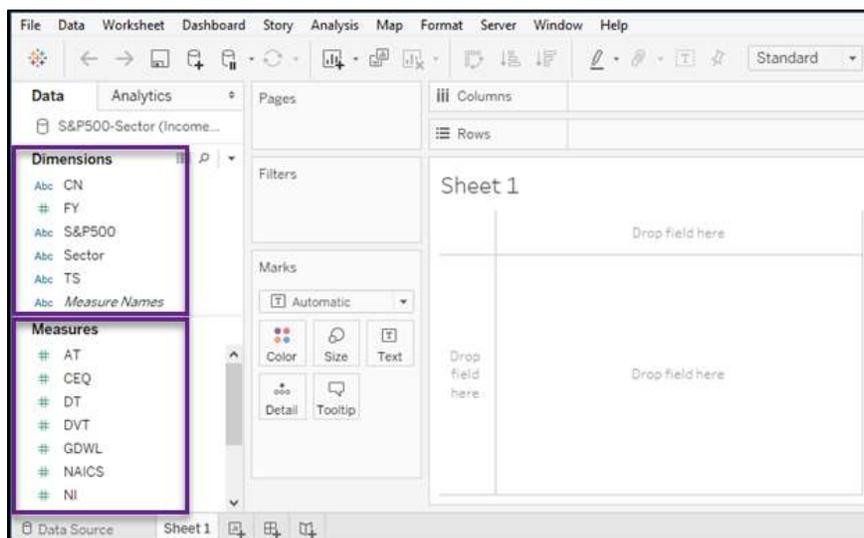


Step 2: Check data types of each data field

Select ‘Sheet 1’ worksheet. Your data fields are grouped into two areas: ‘Dimensions’ and ‘Measures’. The ‘Dimensions’ area contain qualitative values to categorize and the ‘Measures’ area contain quantitative values which can be aggregated (Figure 17).

Figure 17

Dimensions and measures



Drag 'SEC' and 'NAICS' fields from the Measures area to the Dimensions area. The 'SEC' field is a number symbol for the U.S. stock market exchanges and the 'NAICS' is an industry classification code. Even if 'SEC' and 'NAICS' fields show numeric data, the fields should be classified as non-numeric, categorical data. Change data type of 'SEC' and 'NACIS' from number to string (Figure 18).

Figure 18
Tableau data fields in dimensions and measures

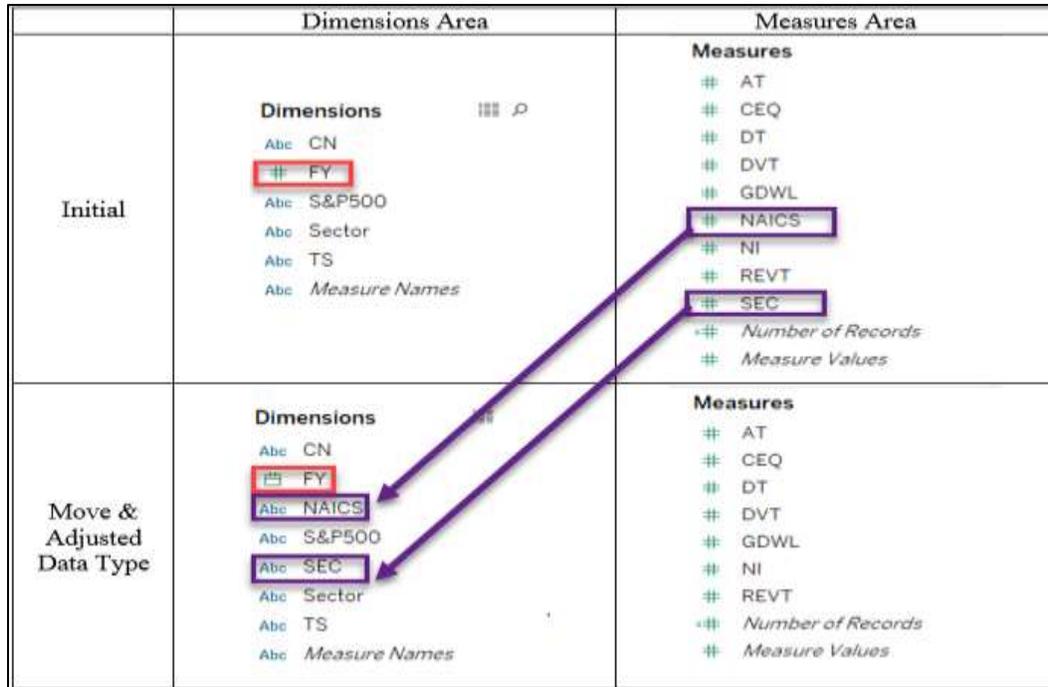
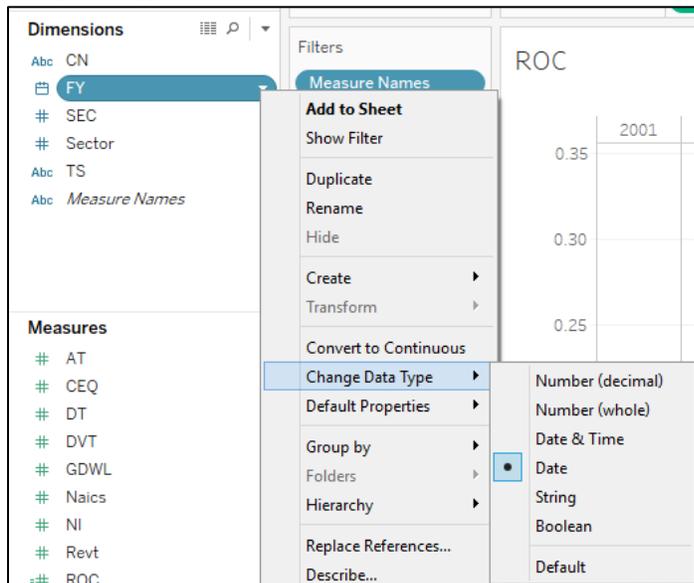


Tableau has four data types: 'String', 'Number', 'Boolean' and 'Date'. Can you find any field in the Dimensions area with an inappropriate data type? Which field should be a 'Date' type? To change a field's data type, right-click the FY field and choose "Change Data Type" menu and choose is 'Date' type (Figure 19).

Figure 19
Change data type

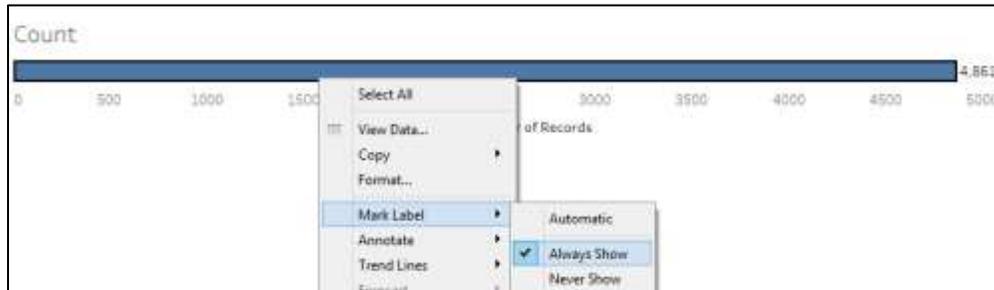


Q13 Check the total number of record and see whether it is the same as the total number of records in ‘S&P500-Sector’ Excel worksheet. Summarize the count of records by year as you did in the Q1.

Drag ‘Number of Records’ in the measures area to ‘Columns Shelf’. Right click the blue bar and select ‘Mark Label’ and ‘Always Show’ (Figure 20) The total number of records is 4,861 (without field header). Rename the worksheet as ‘Count’.

Figure 20

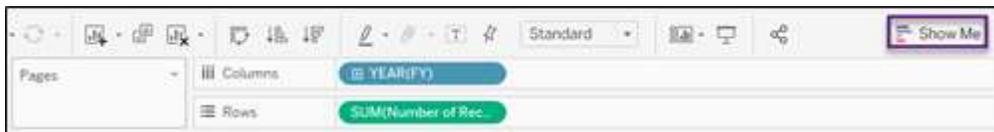
Number of records



Move the field ‘Number of Records’ to ‘Row Self’ and drag ‘FY’ field to ‘Column Shelf’ (Figure 21).

Figure 21

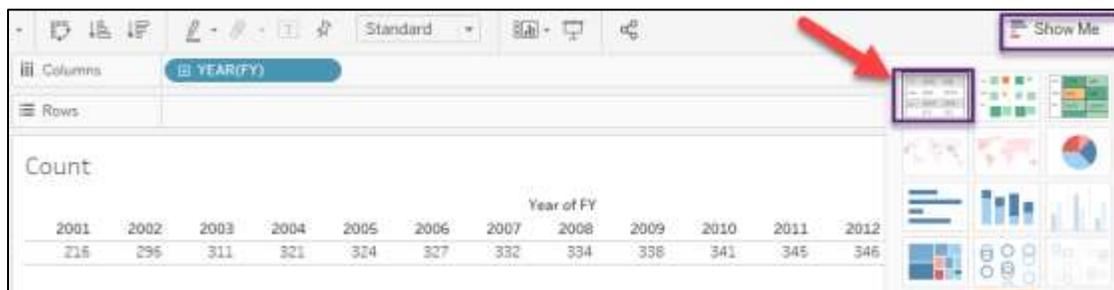
Columns and rows shelves



Click ‘Show Me’ and select ‘text tables’, which is in the upper left corner (Figure 22).

Figure 22

Show me and text table

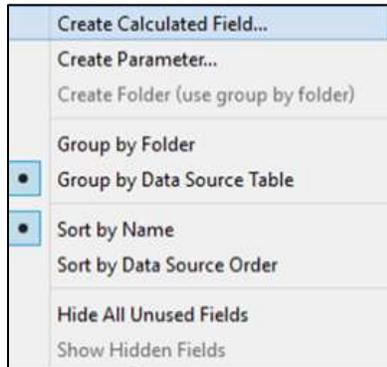


Now you can see a table summarized by year. Click ‘Show Me’ one more time and then the chart options will disappear. Click the ‘New Worksheet’ icon under the bottom of Tableau workbook and rename it as ‘ROC’.

Step 3: Create an ROC field, ROC percentile fields and ROC-Gap field

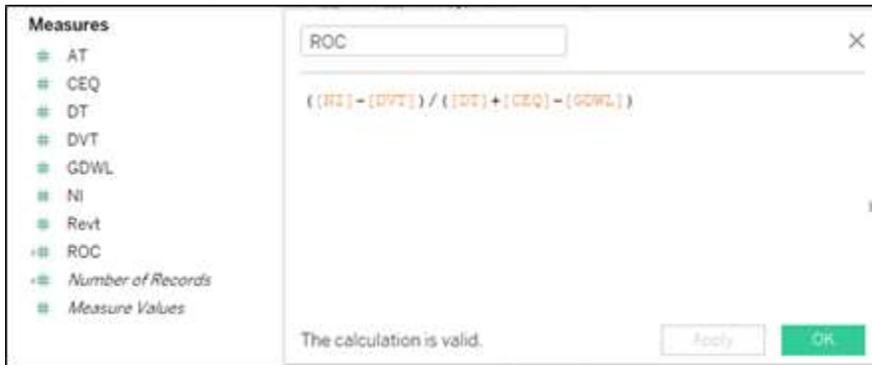
ROC is a ratio indicating how effective a firm is at running invested capital into profits. Profit in the ROC ratio can be net income (NI) or net operating income (NOP). The general equation for ROC is known as $(NI - DVT) / (DT + Equity)$. According to the WSJ article, the ROC in this project is calculated using invested capital without GDWL. Therefore, the ROC formula here is $(NI - DVT) / (DT + CEQ - GDWL)$. Right-click on the white space in Measures area and select ‘Create Calculated Field ...’ (Figure 23) and type the ROC formula.

Figure 23
Create calculated field



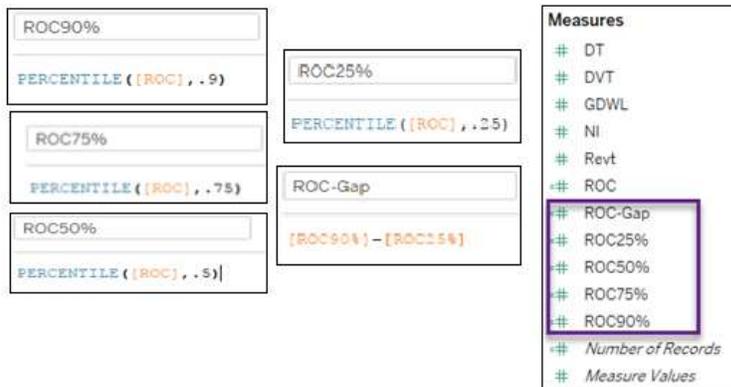
Make sure to use brackets '[']' around data fields (Figure 24).

Figure 24
ROC formula showing use of brackets



In the ascending ordered data set, 90 percentile means a specific number where 90 percentage of data fall below that number. You will now create several more calculated fields: ROC90%, ROC75%, ROC50%, ROC25%, and ROC-Gap, which are the 90%, 75%, 50% and 25% numbers in the ordered ROC values (the ROC-Gap field is 'ROC 90% - ROC 25%'). Use the same method you used to calculate the ROC field. 'Percentile' is a function to return a value which the target percent of field, i.e. percentile ([ROC], .9) returns a value of 90% of ROC. The following is a list of formulas for the above five new calculated fields (Figure 25) Once you create the five new fields, you can see them in the Measures area.

Figure 25
ROCx% and ROC-gap fields



Step 4: Visualize ROC percentiles and ROC-Gap.

Q14 What is the pattern of ROC90%, ROC75%, ROC50%, ROC25% from 2001 through 2015? Increasing, decreasing, or combined of both?

Click ‘New Worksheet’ icon and drag ‘FY’ field to ‘Columns Shelf’ and the four ROC percentiles to ‘Rows Shelf’. Click ‘Show Me’ and choose the ‘Side-by-Side Bars’ chart and click the drop-down arrow of fit icon in the tool bar and choose ‘Entire View’ (Figure 26).

Figure 26

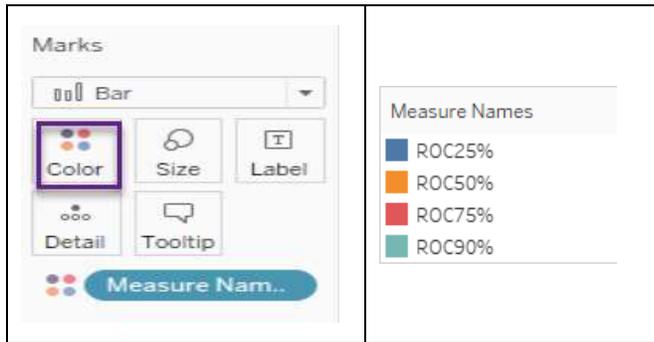
Show me and side-by-side bars



Drag ‘Measure Names’ on the ‘Color’ button in ‘Marks’ card (Figure 27). Right-click on the tab of the worksheet and rename it as ‘ROC’.

Figure 27

Color in marks card and legend

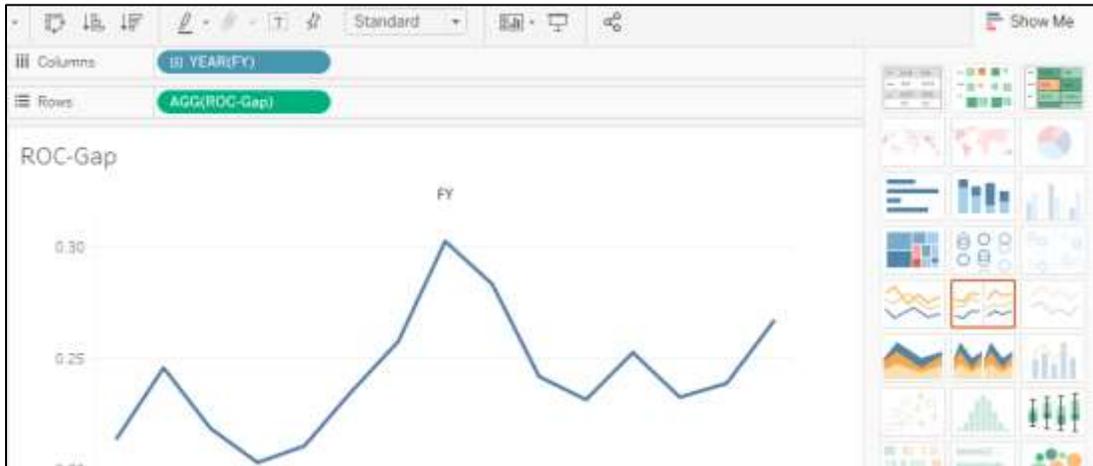


Q15 What is the pattern of ROC-Gap from 2001 through 2015, Increasing, decreasing, or ups and downs? Is it the same pattern as the WSJ article demonstrated? Which year has the largest gap? What would be a reason for the largest gap year?

Drag ‘FY’ field to ‘Column Shelf’ and ROC-Gap to ‘Rows Shelf’. Click ‘Show Me’ and choose the ‘Lines’ chart (Figure 28).

Figure 28

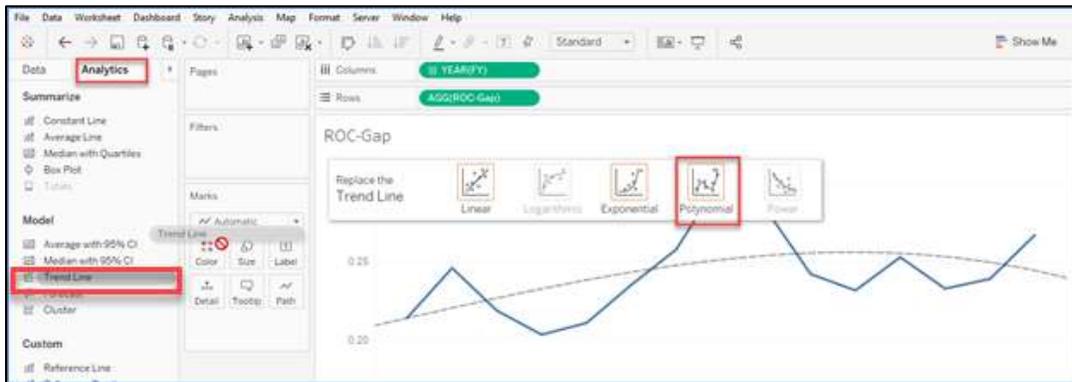
Show me and (discrete) line chart option



Add a trend line using '(Polynomial) Trend Line' in 'Analytics Pane' then right-click on the worksheet tab and rename it 'ROC-Gap' (Figure 29).

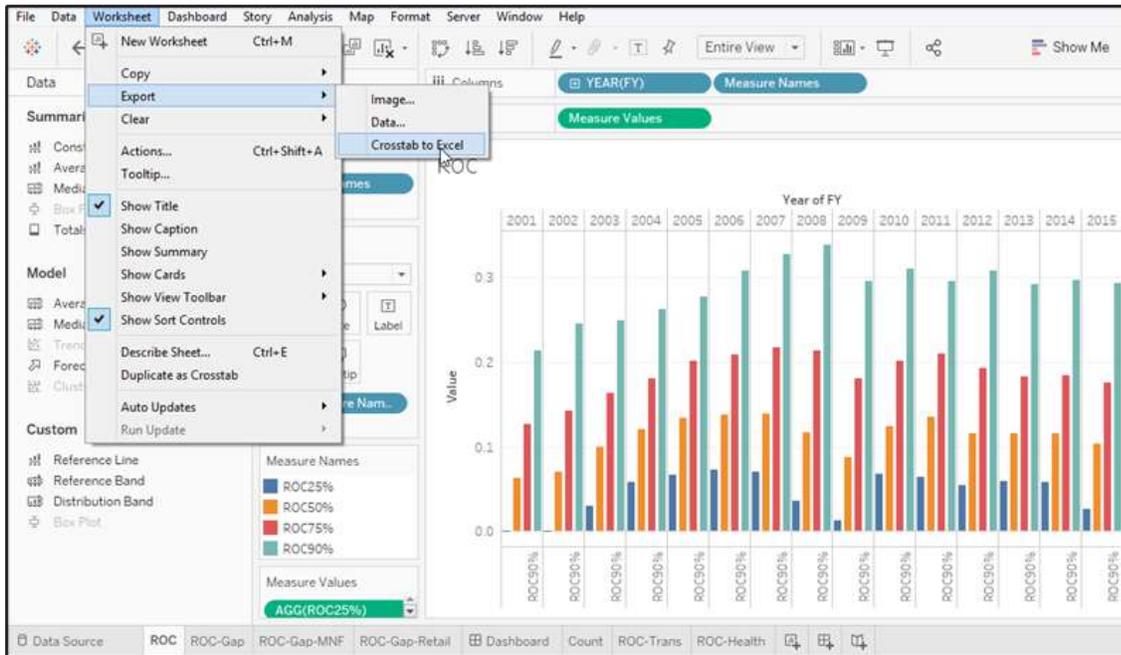
Figure 29

Analytics pane and trend line with polynomial



Q16 Export the worksheet 'ROC' using a menu path of 'Worksheet' / 'Export' / 'Crosstab to Excel' (Figure 30).

Figure 30
Worksheet/export/crosstab to Excel



In the downloaded Excel worksheet (Figure 31), calculate again ROC-Gap at the bottom of the table (row 7) using a formula of 'ROC90% - ROC25%'. Present a combo chart with ROC percentile data (clustered column) and ROC-Gap (line) and compare it with charts generated by Tableau. Copy the downloaded Excel file in the 'Income-Inequality-Student-Data' and right-click on the worksheet tab and rename it as 'ROC-Export'.

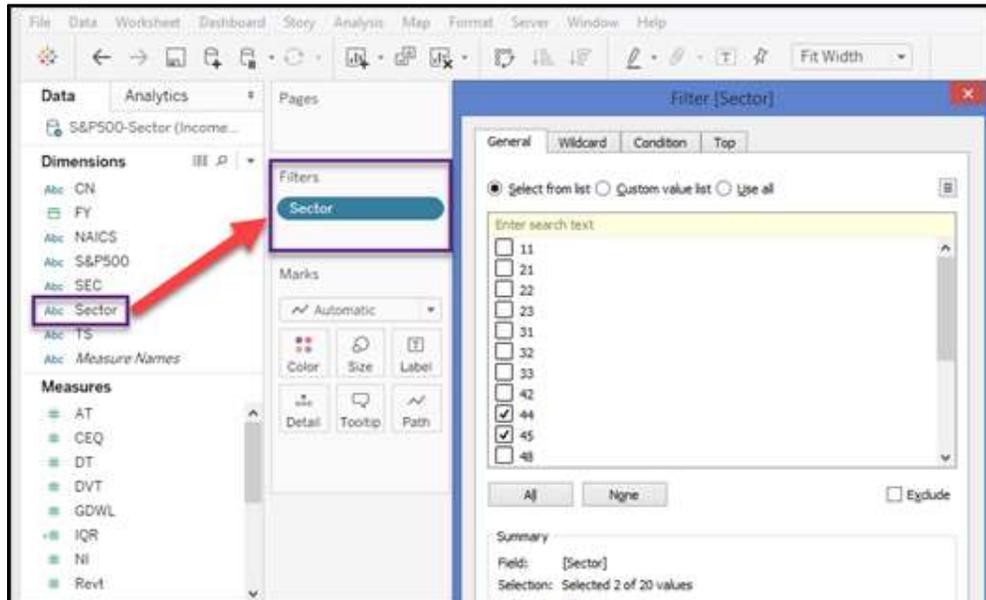
Figure 31
Downloaded "ROC-Gap" Excel file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	FY															
2		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
3	ROC25%	-0.0007	-0.0008	0.0304	0.0587	0.0664	0.0729	0.0599	0.0359	0.0124	0.0677	0.0644	0.0547	0.0594	0.0576	0.0263
4	ROC50%	0.0626	0.0699	0.0993	0.1204	0.1334	0.1383	0.1388	0.1172	0.0874	0.1241	0.1354	0.1162	0.1152	0.1155	0.1039
5	ROC75%	0.1254	0.1429	0.1639	0.1807	0.2018	0.2081	0.2176	0.2141	0.1811	0.2014	0.2100	0.1922	0.1836	0.1839	0.1762
6	ROC90%	0.2135	0.2450	0.2491	0.2621	0.2773	0.3079	0.3276	0.3386	0.2959	0.3067	0.2960	0.3075	0.2921	0.2966	0.2932
7																

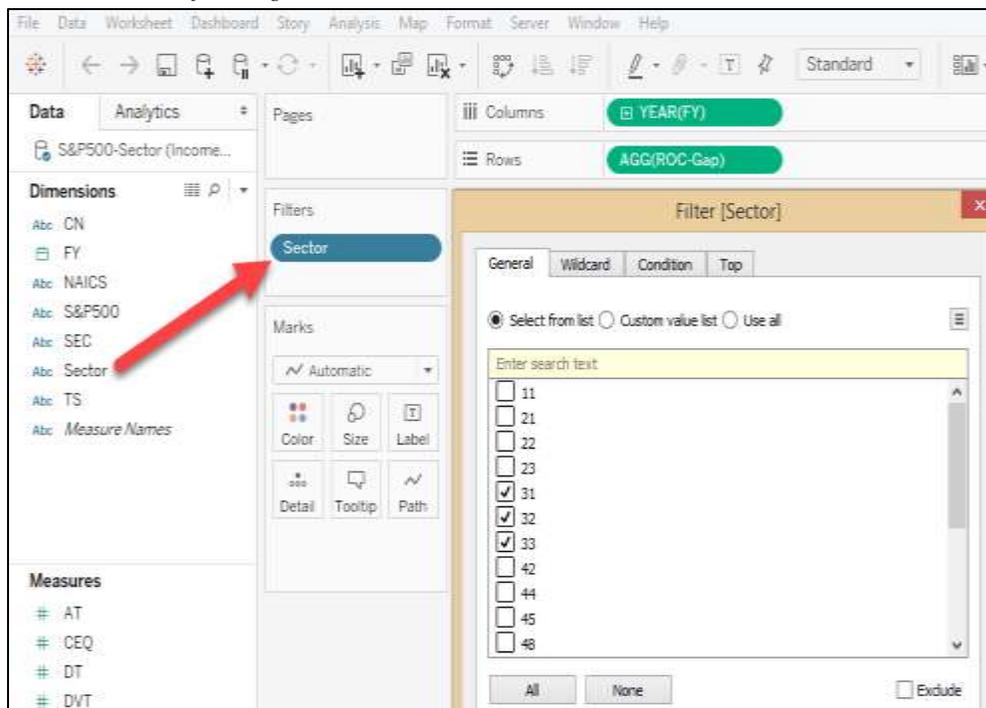
Step 5. Drill-down Analysis

Drill-down is to go deeper into more specific layers of the data being analyzed. In this project, one example of drill-down is to explore data in industry sector level. According to Table 4, sector codes of retail are 44 and 45 and those of manufacturing are 31, 32, and 33.

Q17 Create a chart for ROC-Gap of Retail industry only. Duplicate 'ROC-Gap' worksheet and using Tableau 'Filter' card, extract only retail trade sector (Sector No. = 44, 45). Rename it as 'ROC-Gap-Retail' (Figure 32).

Figure 32*Drill-down to retail trade sector*

Q18 Create a chart for ROC-Gap of manufacturing industry only. Duplicate 'ROC-Gap' worksheet and using Tableau 'Filter' card, extract only manufacturing sectors (Sector No. = 31, 32, 33). Rename it as 'ROC-Gap-MNF'. Comparing ROC-Gap charts, which industry shows a growing significant gap (Figure 33)?

Figure 33*Drill-down to manufacturing sector*

Q19 Export 'ROC-Gap-MNF' and 'ROC-Gap-Retail' to Excel files and make one Excel line chart with polynomial trend lines. You need to combine the two downloaded files into one file to create a chart with data from both manufacturing and retail sectors. Copy the downloaded Excel files in the 'Income-Inequality-Student-Data' and right-click on the worksheet tab and rename it as 'ROC-Sector-Export'.

Q20 Make a dashboard with the four previous worksheets (ROC, ROC-Gap, ROC-Gap-MNF, ROC-Gap-Retail). Click 'New dashboard' icon and drag the four worksheets in the left pane to dashboard area.

Part 4. Technical Memo for Adoption of Tableau

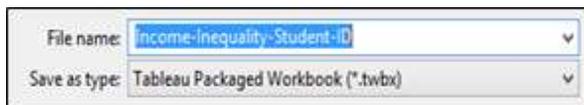
Q21 Write a short e-mail to your CEO on why you need to adopt Tableau assuming your CEO requests a comparison between MS Excel and Tableau. List at least five advantages/disadvantages of Tableau compared with MS Excel, i.e., (1) Visualization, (2) Data Compatibility, (3) Training Efficiency, (4) Cost, and (5) Presentation. Make a new worksheet 'Memo' in 'Income-Inequality-Student-Data.xlsx'

Part 5. Submission

Before you submit your project files, save your Excel and Tableau files with your school student id, i.e., 'Income-Inequality-(Student-ID).xlsx' and 'Income-Inequality-(Student-ID).twbx'. You need to save your tableau file with an extension of "twbx" (Figure 34) The Tableau file with 'twbx' extension is a package of files with your data and your Tableau workbook files for sharing them with your instructor.

Figure 34

Save as twbx extension



Make sure to submit all 16 worksheets, as follows:

Six worksheets in the Tableau file:

1. Count
2. ROC
3. ROC-Gap
4. ROC-Gap-Retail
5. Roc-Gap-MNF
6. Dashboard

Ten worksheets in the Excel file:

1. Data
2. Count
3. S&P500
4. Data-S&P500
5. S&P500-Only
6. S&P500-Not-Null
7. S&P500-Sector
8. Sector
9. Export-ROC
10. Export-Sector-ROC
11. Memo



***AIS Educator Journal* Editorial Board**

Editors-in-Chief

Kimberly Swanson Church, University of Missouri – Kansas City
Gary P. Schneider, California State University, Monterey Bay

Associate Editors

Del DeVries, Belmont University
Dawna Drum, Western Washington University
Betsy Haywood-Sullivan, Rider University
Gail Hoover King, Washburn University
Lorraine Lee, University of North Carolina Wilmington
Conni Lehmann, University of Houston – Clear Lake
Brad Schafer, Kennesaw State University

Senior Reviewers

Kel-Ann Eyler, Georgia College & State University
Kurt Fanning, Grand Valley State University
Cynthia Frownfelter-Lohrke, Samford University
Sonia Gantman, Bentley University
Margaret (Peggy) Garnsey, Siena College
Bonnie Klamm, North Dakota State University
Marcia Watson, University of North Carolina Charlotte
Skip White, University of Delaware

Past Editors-in-Chief

2004-2007 Arlene Savage
2007-2009 Stacy Kovar
2009-2012 David R. Fordham
2012-2015 William G. Heninger
2016-2018 Ronald J. Daigle and David C. Hayes
2018-2019 Chelley M. Vician
2019-2020 Chelley M. Vician and Gary P. Schneider

Editorial Assistant

Abby Bensen, University of St. Thomas

Ad Hoc Reviewers

A list of ad hoc reviewers for the most recent three years is published in the annual editor report.

All materials contained herein are copyright AIS Educator Association, all rights reserved. Faculty members may reproduce any contents of the AIS Educator Journal for use in individual courses of instruction only if the source and the AIS Educator Association copyright are included in any such reproduction. Apply in writing to the Editor for permission to reproduce any AIS Educator Journal content for other uses, including, but not limited to, publication in textbooks and books of readings for general distribution.

The [AIS Educator Journal](#) is published by the [AIS Educator Association](#):

President: Ann O'Brien, University of Wisconsin
Vice President and President Elect: Cynthia Frownfelter-Lohrke, Samford University
Secretary: Cheryl L. Dunn, Grand Valley State University
Treasurer: Kristian Mortenson, University of St. Thomas
Past-President: Dawna Drum, Western Washington University