

## **Hope IT - Data Analysis course - Instructors Guide - February 2021**

This document is meant of instructors who would like to mimic or duplicate the Hope IT Data Analysis course. We felt it worked pretty well, even in an on-line format.

Overall, we had more material than we could cover on 4 Saturdays. It took longer to explain these concepts to those new to data, and data analysis. Our goal is to build confidence in new skills, and reach comprehension, so we are OK with going slow. It was really good to co-teach, as this topic is so large, it takes different perspectives to highlight key topics.

The document covers the following:

- 1) Google Data Studio - Why did we pick this tool, and what were the highlights and challenges.
- 2) Extra topics in Statistics - We want to give our students a bigger picture of how IT analyzes data.
- 3) Scriptural Reflection - Why we picked the Apostle's Creed for this course.
- 4) Datasets and Dashboards: We share the links to our dashboards, as well as the datasets from Kaggle, and our Google Sheets.

We also provide further details in 3 Apendices:

- 1) Appendix A: Start up screens for Data Studio
- 2) Appendix B: Links of Articles and Videos for Data Studio
- 3) Appendix C: Dirty Trick - Y-axis does not start at Zero:

---

### **Google Data Studio:**

- 1) Data Studio uses a Google Account, which requires students to be 13 or over. We targeted this for high schoolers. 13 is about the right age anyways, as the ideas and skill level are around the age of 13 or higher.
- 2) Data Studio is free, and on-line. And we used a sharing feature in Data Studio that a student can give a link to others. This was great way for us to see the student's dashboard while they were changing it. This worked out great, and for Week 3 and 4 in this course, we used the same link for both weeks.
- 3) We show the screen shots one will see when first starting, see Appendix A below.
- 4) Data Studio was a good tool to use, both for basic statistical functions, like minimum, maximum, count, sum, mean (average), and median. It was very good to go over this with the students each week.
- 5) The charts worked out great. We went from Bar, to Line, to Geo charts.
- 6) Data Blending was the hardest topic to go through. It was both a different idea, if a student isn't used to databases (which most are not.) It is also difficult to walk student through; it took us 2 times to get it right. We planned to blend 4 datasets together, but ended up just blending 2 datasets, and we merged 3 of our Google sheets together, the

StateNames, the ISO3166 values, and the 2019Population. Here is where we got this info:

- a) ISO3166 value: <https://www.iso.org/obp/ui/#iso:code:3166:US>
  - b) 2019 Population:  
<https://usafacts.org/data/topics/people-society/population-and-demographics/population-data/population/>
- 7) If you want to know the articles and videos we used to learn about Data Studio, we list them in Appendix B. These may help you as well.

#### Extra topics on Statistics:

- 1) We covered some dirty tricks used to make charts or surveys seem to suggest one thing, but really not be true. We used ideas from the book, [How to Lie with Statistics](#), by Darrell Huff.
  - a) Data sampling not covering a good sample of the total population will lead to bad conclusions.
  - b) Not starting your charts at zero will over-emphasize changes from one measurement to the next, for example growth can look larger than it really is, fooling a customer. This is shown in Data Studio in more detail in Appendix C below.
- 2) 4 quadrants ([Anscombe's Quartet](#), [Jupyter Notebook on the Quartet](#)): These 4 datasets have nearly the same mean (average) for their x and y values, and other statistical metrics. But, they are very different. This is a cautionary example to make sure one looks at your dataset carefully, and for sure, create a few charts.
- 3) Predicting future events with Data Science, Models and Statistics. We are very fortunate to have a team member that can speak on these topics with some experience.

#### **Scriptural Reflection:**

Hope IT is Christian-based. We not only want to teach tech skills, but also allow students to gain a better understanding of how God has been in their past, walk with them in the present, and intends a good future for them. Reading Scripture in our sessions is one great way of doing that.

We picked reviewing the Apostle's Creed for the following reasons:

- 1) In data, we stand on facts. Facts create our core beliefs on a topic. In our Christian faith, what do we stand on? We believe core understandings, and these are best expressed in some of our creeds. The Apostle's Creed was finalized around 390 AD. That's a long time ago. It has stood the test of time, and many denominations believe this, including the Catholic Church. Below are 2 links showing this:
  - a) Catholic Church: Franciscan Mission Associates:  
[https://franciscanmissionassoc.org/rosarybook\\_3/?gclid=Cj0KCQiAkuP9BRCKARIsAKGLE8V9Fa6sAe72LoQEEdm21nbEFenL7ESj4ncGlxw183fmZv7Na4YRT3V0aAIPmEALw\\_wcB](https://franciscanmissionassoc.org/rosarybook_3/?gclid=Cj0KCQiAkuP9BRCKARIsAKGLE8V9Fa6sAe72LoQEEdm21nbEFenL7ESj4ncGlxw183fmZv7Na4YRT3V0aAIPmEALw_wcB)
  - b) Book of Confession (USA Presbyterian):  
<https://www.pcusa.org/resource/book-confessions/>

- 2) In February, we are in the middle between Christmas and Lent/Easter. We covered the 3 lines in the Apostle's Creed dealing with Christ's birth, death, and resurrection.
- 3) We used the Gospel of Luke to cover the details. Luke was a physician, so he was detailed oriented.

### **Datasets and Dashboards:**

We got our datasets from Kaggle. Links for those are below, as well as the dashboards we created.

- 1) The Cell Phone of members in the course was a great way to start our class. We started slow, explaining Rows and Columns/Fields, and Data Types. We used our personal data to show how data is collected, and that data represents real facts.
- 2) The Video Game Sales dataset was large, and made the Google Sheet quite large. We were reducing this to the first 400 (top 400 by rank), for some kids to download, but it turned out they didn't show up for that week, so we didn't use the shortened version. We knew their mins, maxs, and other statistical functions would come up different than those using the full dataset.
- 3) For COVID, we have to wrangle/clean it up a bit. We had to reduce it as it was too large for Google sheets, and too large would be too large for our kids to upload. We got rid of all the counties, and summed them up as states. As the counts are cumulative, we also had to subtract the next day from the previous. This led to some curious results, where some values (about 44 of them) were negative. It looked odd when we did our minimum checks to find we have one State for one day to have -403 cases. How could that be? We had to explain.
- 4) We were hoping to walk with our students how to get into Kaggle, clean/wrangle data, and then get it into a Google sheet. We just didn't have enough time, but we think that will be a very important exercise.

Below are the links we used to get data for our datasets:

#### **Cell Phone samples from our Hope IT course members:**

DataStudio:

<https://datastudio.google.com/reporting/464ee41e-0984-49bb-897b-5ced12124648>

Google Sheet:

<https://docs.google.com/spreadsheets/d/1knM0Eq-ZLf08HN0dXSMPInVI636GtsQaeaBhiSQXZe8/edit?usp=sharing>

#### **Video Game Sales:**

Kaggle dataset:

<https://www.kaggle.com/gregorut/videogamesales>

DataStudio:

<https://datastudio.google.com/reporting/6a5b2260-5fab-40b3-b631-d757c2188e98>

Google Sheet:

[https://docs.google.com/spreadsheets/d/1Fnqa6lC5PWc\\_xd9KfJDGJJi\\_MrJzNmFaU8l6qX\\_3hxU/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1Fnqa6lC5PWc_xd9KfJDGJJi_MrJzNmFaU8l6qX_3hxU/edit?usp=sharing)

Shortened Google Sheet (top 471 by Global Sales):

<https://docs.google.com/spreadsheets/d/1WLjdhMDSWlofLncJa5jRCnn-ouPE7Kxfk3qcqOrbVmo/edit#gid=2045852336>

Ideas of questions on the Video Game Sales data:

<https://www.kaggle.com/snanilim/video-games-sales-analysis-and-visualization>

### **COVID-USA-Confirmed Cases:**

Kaggle dataset:

<https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university>

2019 USA State Population dataset:

<https://usafacts.org/data/topics/people-society/population-and-demographics/population-data/population/>

Data Studio:

<https://datastudio.google.com/reporting/4f85fffd-8b50-463c-a7e6-da877fce86b0>

COVID Google Sheet (up to Feb. 18, 2021):

[https://docs.google.com/spreadsheets/d/1Fsk\\_AGIMSkLpQBxz2BC4s-9\\_SQqMOW4KSfPGbRml\\_3g/edit#gid=697306565](https://docs.google.com/spreadsheets/d/1Fsk_AGIMSkLpQBxz2BC4s-9_SQqMOW4KSfPGbRml_3g/edit#gid=697306565)

Combined StateName, ISO3166 and 2019 Population Google Sheet:

<https://docs.google.com/spreadsheets/d/1hYB65uW4RyH6UjubiiikLM2Tah5idfrOJLq1kgJ8mA/edit#gid=1629225043>

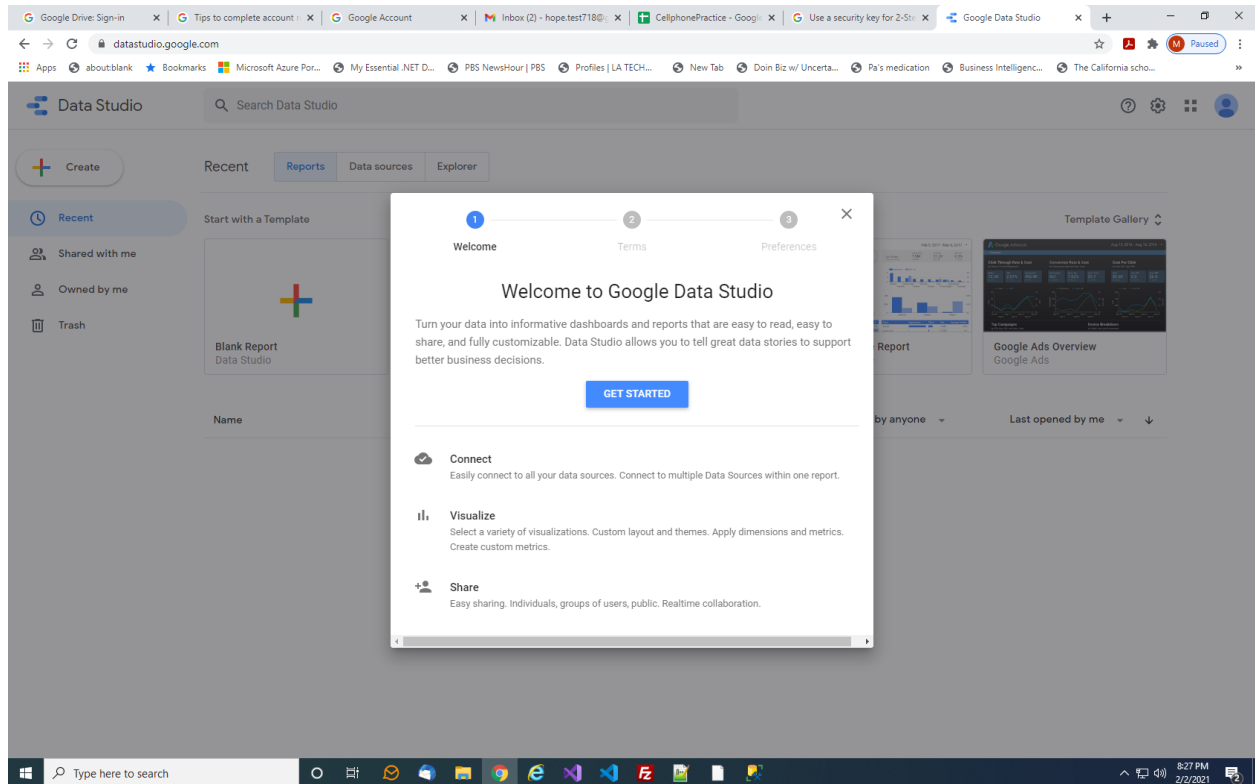
Ideas of questions on COVID data (US and Global):

<https://www.kaggle.com/therealcyberlord/coronavirus-covid-19-visualization-prediction>

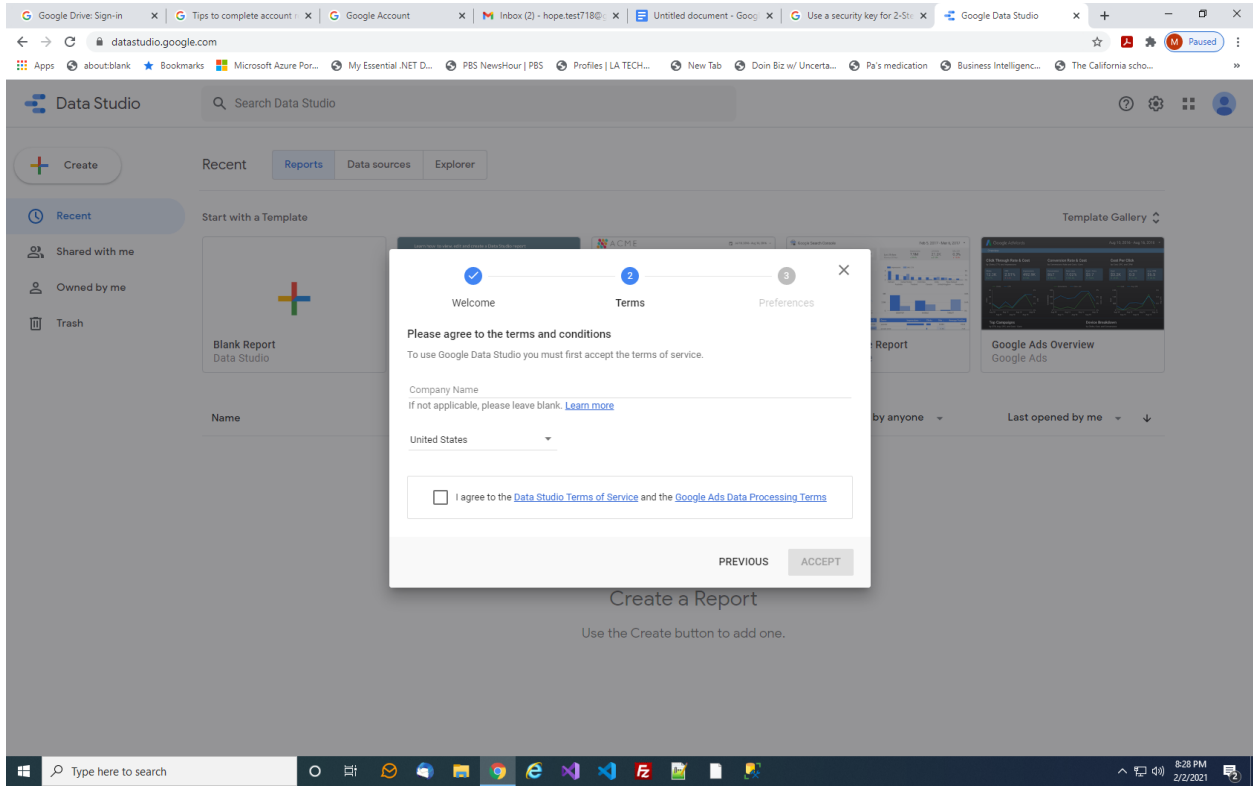
## Appendix A - Startup screen shots:

The following screen shots are what a new users will see when starting Data Studio. These are the Data Studio Terms and Agreements. The last 2 screen shots are what users will see when connecting to Google Sheets, and some popups and selections of the Google sheet that are needed.

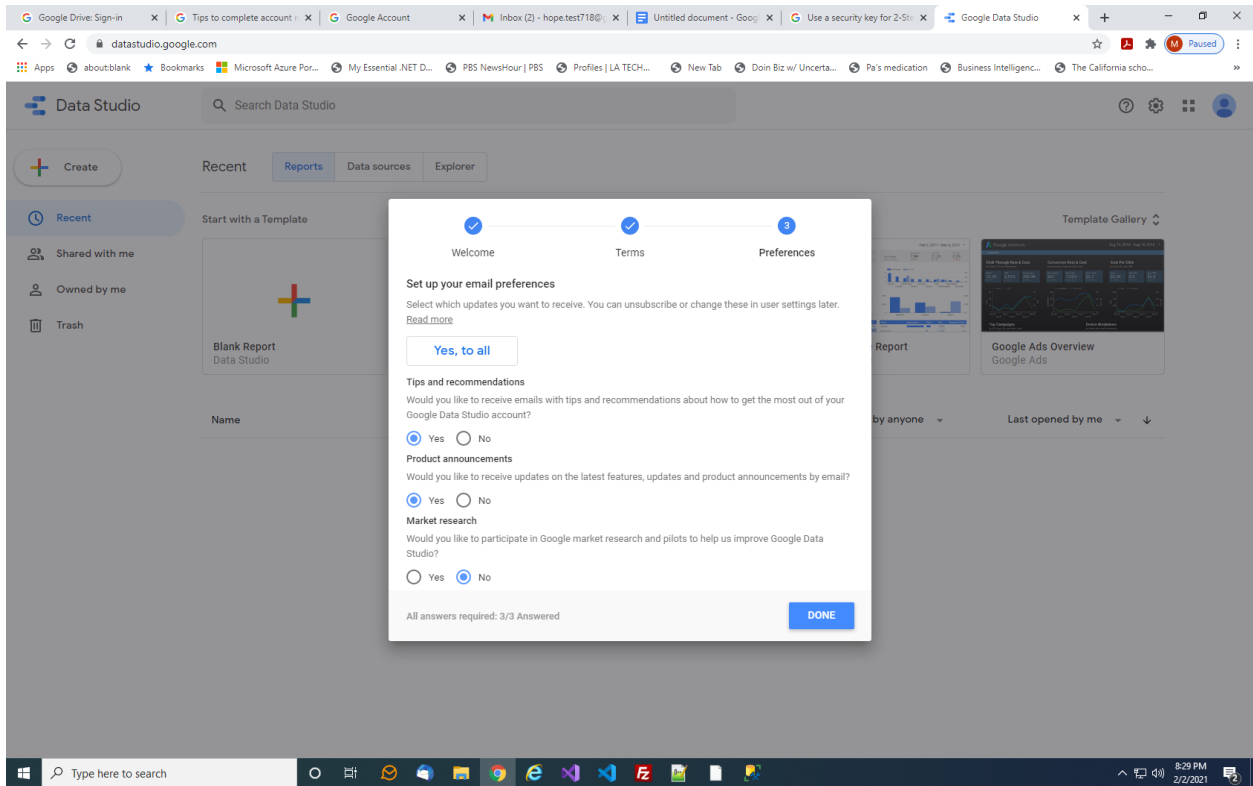
#1



#2



#3



## #4: Select Google Sheets

The screenshot shows the Data Studio interface. At the top, there's a browser window with multiple tabs. The main window displays an 'Untitled Report' with a menu bar (File, View, Page, Help) and a toolbar with options like 'Add a page', 'Add data', 'Add a chart', and 'Add a control'. Below the report area, a dialog box titled 'Add data to report' is open. It has a search bar and a section for 'Google Connectors (18)'. The connectors are arranged in a grid:

- Google Analytics**: By Google. Connect to Google Analytics.
- Google Ads**: By Google. Connect to Google Ads performance report data.
- Google Sheets**: By Google. Connect to Google Sheets.
- BigQuery**: By Google. Connect to BigQuery tables and custom queries.
- File Upload**: By Google. Connect to CSV (comma-separated values) files.
- Campaign Manager 360**: By Google. Connect to Campaign Manager 360 data.
- Cloud Spanner**: By Google. Connect to Google Cloud Spanner databases.
- Cloud SQL for MySQL**: By Google. Connect to Google Cloud SQL for MySQL databases.
- Display & Video 360**: By Google. Connect to Display & Video 360 report data.
- Extract Data**: By Google. Connect to Extract Data.
- Google Ad Manager 360**: By Google. Connect to Google Ad Manager data.
- Google Cloud Storage**: By Google. See your files in Google Cloud Storage.

The Windows taskbar at the bottom shows the time as 8:31 PM on 2/2/2021.

## #5: Authorize Data Studio to connect/access Google Sheets

Google Drive: Sign-in | Tips to complete account | Google Account | Inbox (2) - hope.test718@ | Untitled document - Goo... | Use a security key for 2-5i... | Untitled Report

datastudio.google.com/reporting/4f80fcbb-f32f-49f8-9b1b-fe102d5e5da6/page/IE3z8/edit

Apps | about:blank | Bookmarks | Microsoft Azure Por... | My Essential .NET D... | PBS NewsHour | PBS | Profiles | LA TECH... | New Tab | Doin Biz w/ Uncerta... | Pa's medication | Business Intelligenc... | The California scho...

Untitled Report  
File View Page Help

Add a page | Add data | Add a chart | Add a control | Theme and layout

← Add data to report Data credentials: Owner

**Google Sheets**  
By Google  
The Google Sheets connector allows you to access data stored in a Google Sheets worksheet.  
[LEARN MORE](#) [REPORT AN ISSUE](#)

Authorization

Data Studio requires authorization to connect to your Google Sheets.

**AUTHORIZE**

Cancel Add

Type here to search 8:31 PM 2/2/2021

## #6: Pick the Google Sheet you would like to use

Google Drive: Sign-in | Tips to complete account | Google Account | Inbox (2) - hope.test718@ | Untitled document - Goo... | Use a security key for 2-5i... | Untitled Report

datastudio.google.com/reporting/4f80fcbb-f32f-49f8-9b1b-fe102d5e5da6/page/IE3z8/edit

Apps | about:blank | Bookmarks | Microsoft Azure Por... | My Essential .NET D... | PBS NewsHour | PBS | Profiles | LA TECH... | New Tab | Doin Biz w/ Uncerta... | Pa's medication | Business Intelligenc... | The California scho...

Untitled Report  
File View Page Help

Add a page | Add data | Add a chart | Add a control | Theme and layout

← Add data to report Data credentials: Owner

**Google Sheets**  
By Google  
The Google Sheets connector allows you to access data stored in a Google Sheets worksheet.  
[LEARN MORE](#) [REPORT AN ISSUE](#)

ALL ITEMS	Spreadsheet
OWNED BY ME	CellphonePractice
SHARED WITH ME	
STARRED	
URL	
OPEN FROM GOOGLE DRIVE	

Cancel Add

Type here to search 8:32 PM 2/2/2021



## **Appendix B - Links of Articles and Videos for Data Studio**

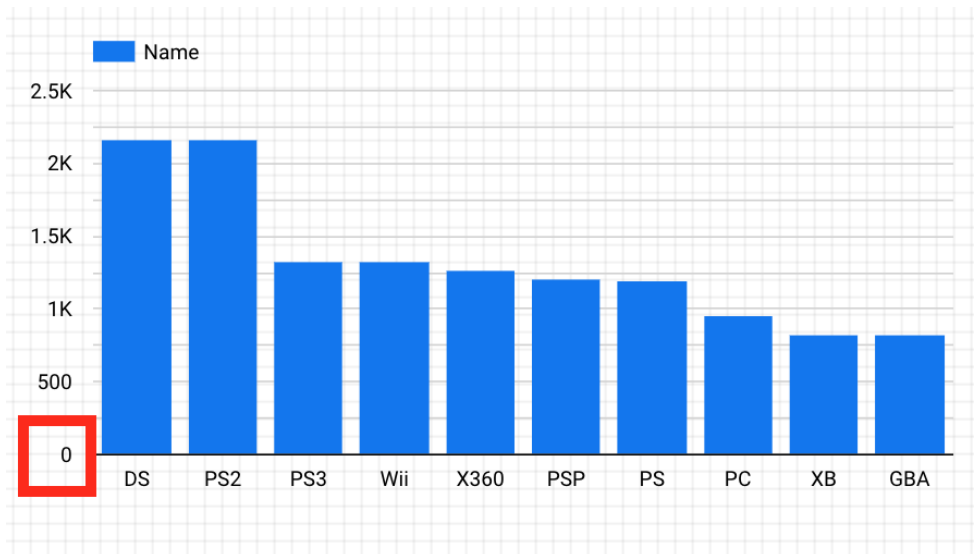
As of February 2021, we used many Internet articles and videos to learn about Data Studio, to try it out ourselves, and be able to teach kids. Below are the best ones we recommend to you to also learn, although there are more articles and videos out there.

- Overall information: <https://blog.hubspot.com/marketing/google-data-studio>
- Overall (video):  
[https://www.youtube.com/watch?v=R0rV4ZS-ruQ&ab\\_channel=MeasureSchool](https://www.youtube.com/watch?v=R0rV4ZS-ruQ&ab_channel=MeasureSchool)
- Connect Google Sheets to Data Studio:  
[https://www.youtube.com/watch?v=EcMN3sl918M&ab\\_channel=MeasurementMarketing%2Cio](https://www.youtube.com/watch?v=EcMN3sl918M&ab_channel=MeasurementMarketing%2Cio)
- Shows you slowly each part of Data Studio:  
[https://www.youtube.com/watch?v=1qGsimmHiu8&ab\\_channel=railsware](https://www.youtube.com/watch?v=1qGsimmHiu8&ab_channel=railsware)
- Sharing your dashboard: <https://support.google.com/datastudio/answer/6287179?hl=en>
- Data Blending1: <https://support.google.com/datastudio/answer/9061420?hl=en>
- Data Blending2:  
[https://www.youtube.com/watch?v=UCAjNKzfdQ&ab\\_channel=MeasureSchool](https://www.youtube.com/watch?v=UCAjNKzfdQ&ab_channel=MeasureSchool)
- Geo charts:  
[https://www.youtube.com/watch?v=ExnLW\\_c3QHA&ab\\_channel=LearnDigitalAdvertising](https://www.youtube.com/watch?v=ExnLW_c3QHA&ab_channel=LearnDigitalAdvertising)
- Calculated fields: <https://support.google.com/datastudio/answer/6299685?hl=en>

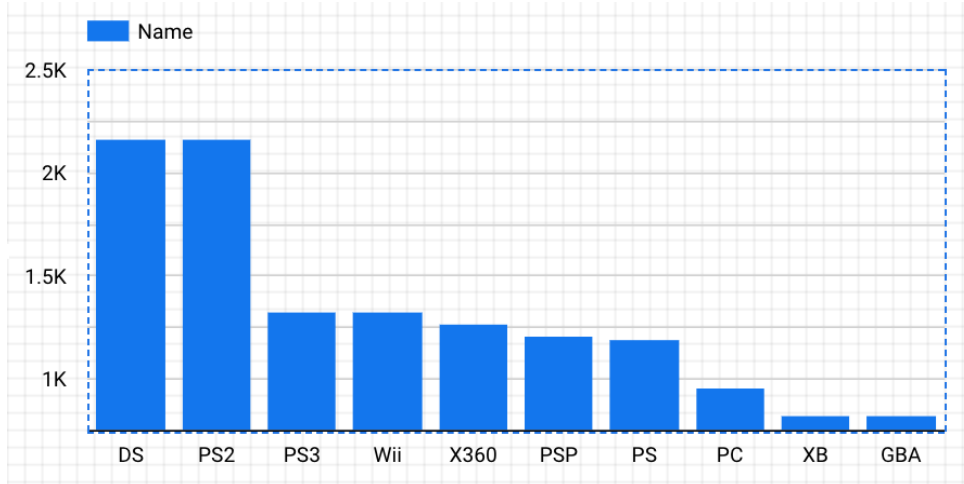
### **Appendix C: Dirty Trick - Y-axis does not start at Zero:**

The following screen shots show what Daniel Huff, in his Chapter 5 “The Gee-Whiz Graph” of his book “How to Lie with Statistics”, shows how we can give the viewer the wrong impression. Sometimes this is intentionally used to shock or deceive the viewer, that the progress is really better than it is. We can demonstrate that in Data Studio, and to be a respected and trust data analyzer, we need to avoid this. We need to learn to not only use facts, but tell the truth about them.

The below shows how a normal bar chart in Data Studio starts out. The below bar chart shows how many game platforms are sold over the period of time for the entire dataset. DS and PS2 are above and beyond the rest. Notice in the bottom left, the Y-Axis starts at zero:



Now, Data Studios allows us to use these same facts, but possible exaggerate how much DS and PS2 are above the rest. Take a look at the below. Does this tell a different story? Same facts, but now DS and PS2 look about 3 times more than the next level of platforms. It makes DS and PS2 sales look a lot more impressive.



The below shows how to change the Y-axis, to make this example work. Go to the “Style” tab of the Chart, and then scroll down to the “Left Y-axis” section, and change the “Axis Min” setting to 750. One should probably never change the Y-axis to be different from zero.

|
|
|
|
Chart > Bar
▼

DATA

STYLE

Axes

Show axes

Reverse Y-axis direction

Reverse X-axis direction

Left Y-Axis

Show axis title

Axis Min

750

↕

Axis Max

(auto)

Custom Tick Interval

(auto)