

Data Jam

Workshop 1: Tableau

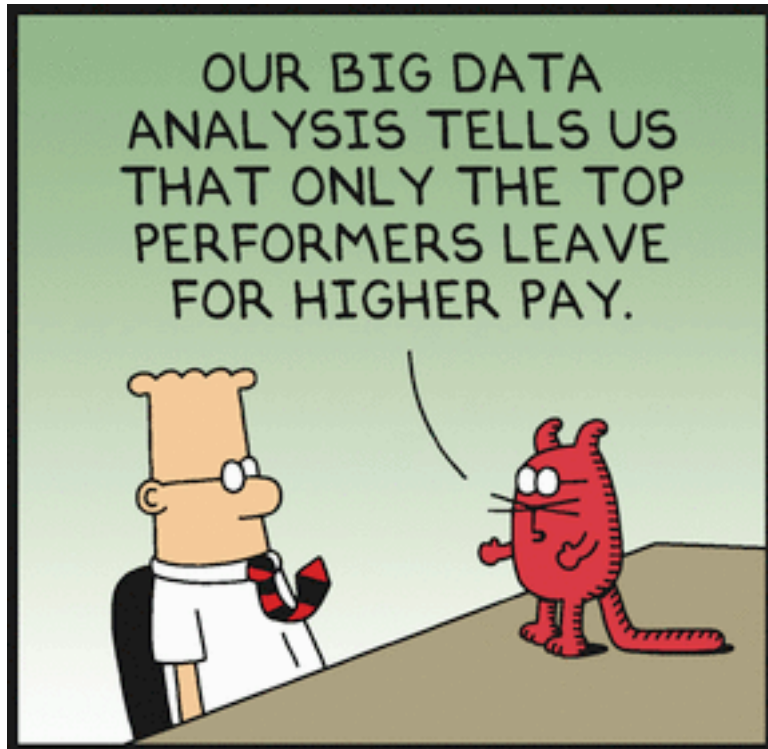
Presenter: Mark Voortman

<https://datajam.it.pointpark.edu/>

(all materials downloadable)



Introduction – What is this Data Jam About?



Dilbert.com DilbertCartoonist@gmail.com



1-27-14 ©2014 Scott Adams, Inc. /Dist. by Universal Uclick



Introduction – What is this Data Jam About?

(Big) Data



HR Attrition
who leaves



Insights



Introduction – Why this Data Jam? It's Fun!

Harvard Business Review:

Data Scientist: The Sexiest Job of the 21st Century

The shortage of data scientists is becoming a serious constraint in some sectors.

Glassdoor: 50 Best Jobs in America for 2019

#1: Data Scientist

\$108k median salary

6,500 job openings

4.3/5 job satisfaction

Logistics – Important Dates

- Workshops
 - February 20th (today) – Tableau (data visualization and exploration)
 - March 6th – Weka (predictive analytics software)
- Poster Competition
 - April 3rd – Poster Presentations (present your results!)
 - More details next workshop



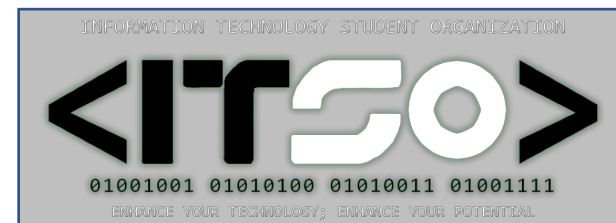
Logistics – The Prizes 😊

- 1st Prize
 - Data Jam Trophy
 - Barnes & Noble gift cards – \$100 per person!
- Runner-up Prize
 - Stage AE tickets
- Everyone Else Wins ...
 - ... new skills!



Logistics – Random Notes

- Team formation
- You can use any tool you want
 - We teach you Tableau and Weka
 - But feel free to use any other tool (Excel, Python, etc.)
- Judges
 - Industry professionals
 - Very experienced with data and modeling
 - Names, titles, and affiliations to be announced
- The Data Jam is co-organized with ITSO: <http://itso.pointpark.edu/>
 - Join ITSO if you like this kind of stuff



Slack – A Tool for Communication

- Slack is a popular **communication** tool used by many **tech** companies
- Go to <https://pointparkuniversity.slack.com/> and join
- Use for
 - Reaching out to mentors with questions
 - Team collaboration
- Apps available for iOS, Android, etc.
- See next slides for steps and screen shots



Slack – Creating an Account

Sign in to Point Park University

pointparkuniversity.slack.com

Enter your email address and password.

 Remember me
[Forgot password?](#) · [Forgot which email you used?](#)

If you have an @pointpark.edu email address, you can [create an account](#).

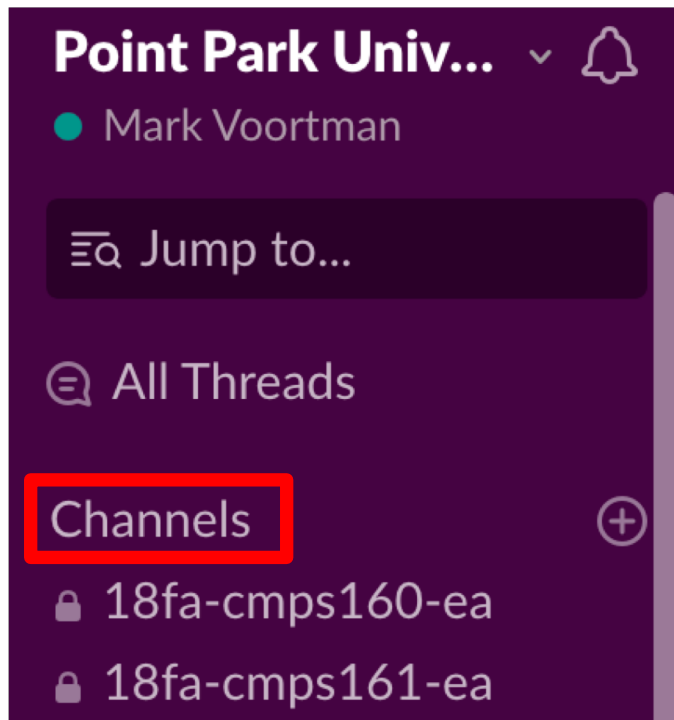
Trying to create a workspace? [Create a new workspace](#)



Sign up for pointparkuniversity.slack.com

Enter your email address to get started.

Slack – Joining the #datajam Channel



Point Park Univ... ▾ 🔔

● Mark Voortman

🔍 Jump to...

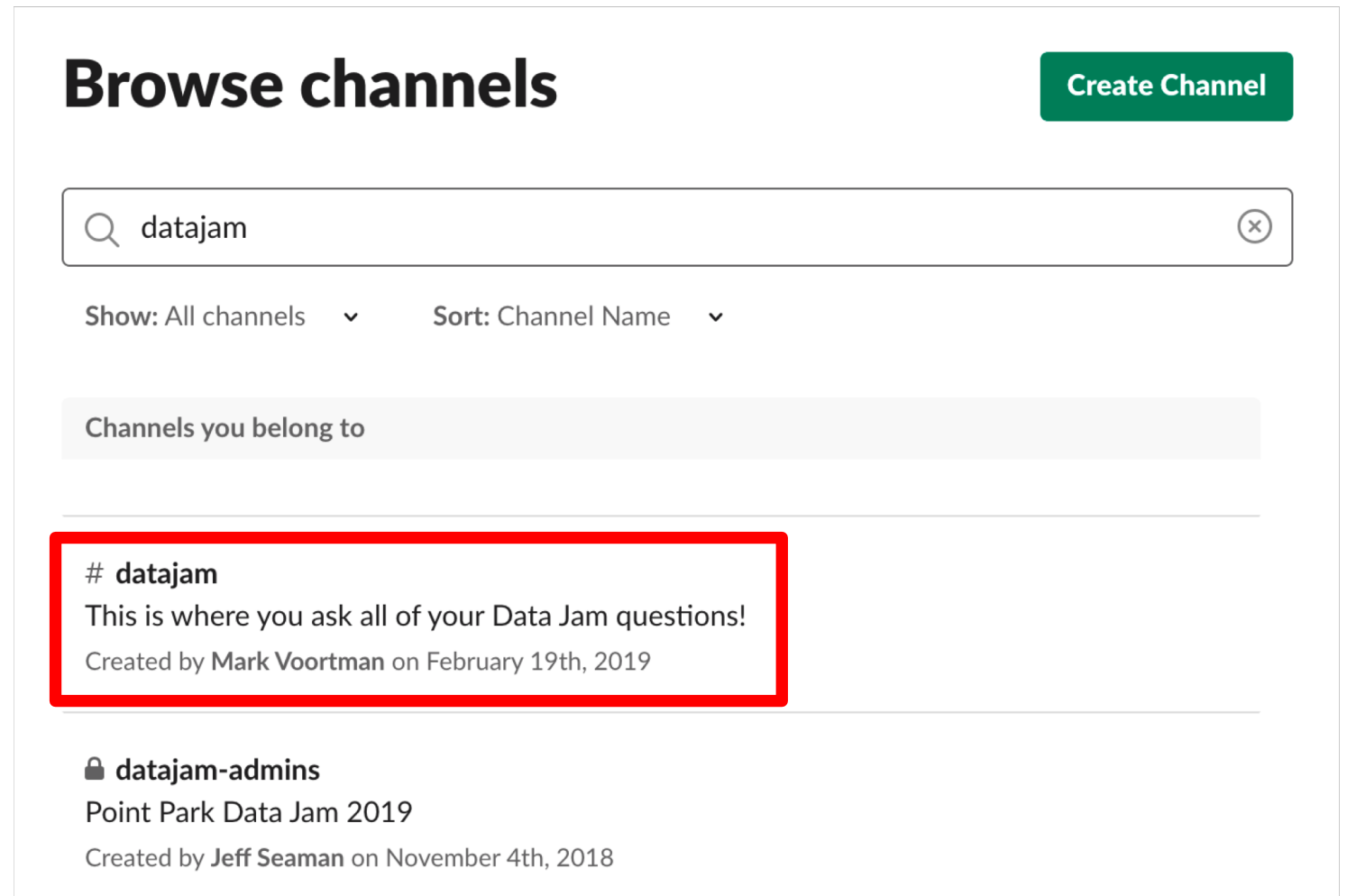
🗨 All Threads

Channels ⊕

🔒 18fa-cmps160-ea

🔒 18fa-cmps161-ea

A blue arrow points from the 'Channels' section to the right.



Browse channels Create Channel

🔍 datajam

Show: All channels ▾ Sort: Channel Name ▾

Channels you belong to

datajam
This is where you ask all of your Data Jam questions!
Created by Mark Voortman on February 19th, 2019

🔒 **datajam-admins**
Point Park Data Jam 2019
Created by Jeff Seaman on November 4th, 2018

Slack – Sending #datajam Messages

Point Park Univ... | Mark Voortman

#datajam | 1 | 0 | Add a topic

Jump to...

All Threads

Channels

- 18fa-cmps160-ea
- 18fa-cmps161-ea
- 18fa-cmps260-ea
- 18fa-cmps260-oaon
- 18fa-cmps361-ea
- 18sp-cmps163-ea
- 18sp-cmps262-oaon
- 18sp-cmps480-ea
- alumni
- class-of-2019
- class-of-2020
- class-of-2021
- class-of-2022
- clubs
- # datajam**
- datajam-admins
- faculty
- # general
- gfk-hackathon
- # itso
- # p-actordatabase
- # p-pioneerpantry
- # p-pointlift
- # p-portfoliobuilder
- # p-ppu-app
- # p-ppu-sh
- # p-ppuesports
- # p-robocodejs
- # p-robothelp
- # p-syllabusbuilder
- # p-why-inf-tech

datajam

You created this channel today. This is the very beginning of the **# datajam** channel. Purpose: This is where you ask all of your Data Jam questions! ([edit](#))

[+ Add an app](#) [🔗 Invite others to this channel](#)

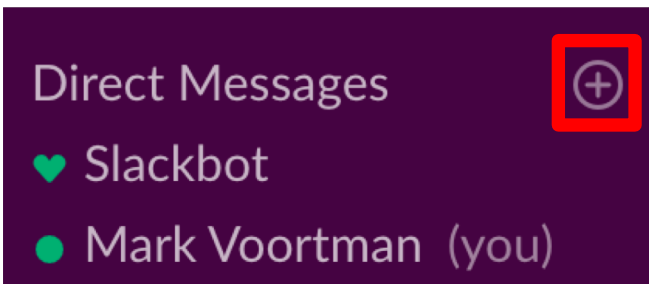
Today

Mark Voortman 8:32 PM
joined #datajam.

Mark Voortman 8:32 PM
set the channel purpose: This is where you ask all of your Data Jam questions!

+ Message #datajam @ 🗨️

Slack – Sending Direct Messages



The Data – How Do I Obtain It?

Download from

<https://datajam.it.pointpark.edu/hr-employee-attribution.csv>

hr-employee-attribution.csv

CSV format

What does that mean?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	Environment	Gender	HourlyRate	JobInvolvement	JobLevel
2	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	Female	94	3	2
3	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	3	Male	61	2	2
4	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4	Male	92	2	1
5	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	4	Female	56	3	1
6	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1	Male	40	3	1
7	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	8	4	Male	79	3	1
8	59	No	Travel_Rarely	1324	Research & Development	3	3	Medical	1	10	3	Female	81	4	1
9	30	No	Travel_Rarely	1358	Research & Development	24	1	Life Sciences	1	11	4	Male	67	3	1
10	38	No	Travel_Frequently	216	Research & Development	23	3	Life Sciences	1	12	4	Male	44	2	3
11	36	No	Travel_Rarely	1299	Research & Development	27	3	Medical	1	13	3	Male	94	3	2
12	35	No	Travel_Rarely	809	Research & Development	16	3	Medical	1	14	1	Male	84	4	1
13	29	No	Travel_Rarely	153	Research & Development	15	2	Life Sciences	1	15	4	Female	49	2	2
14	31	No	Travel_Rarely	670	Research & Development	26	1	Life Sciences	1	16	1	Male	31	3	1
15	34	No	Travel_Rarely	1346	Research & Development	19	2	Medical	1	18	2	Male	93	3	1
16	28	Yes	Travel_Rarely	103	Research & Development	24	3	Life Sciences	1	19	3	Male	50	2	1
17	29	No	Travel_Rarely	1389	Research & Development	21	4	Life Sciences	1	20	2	Female	51	4	3
18	32	No	Travel_Rarely	334	Research & Development	5	2	Life Sciences	1	21	1	Male	80	4	1
19	22	No	Non-Travel	1123	Research & Development	16	2	Medical	1	22	4	Male	96	4	1
20	53	No	Travel_Rarely	1219	Sales	2	4	Life Sciences	1	23	1	Female	78	2	4
21	38	No	Travel_Rarely	371	Research & Development	2	3	Life Sciences	1	24	4	Male	45	3	1
22	24	No	Non-Travel	673	Research & Development	11	2	Other	1	26	1	Female	96	4	2
23	36	Yes	Travel_Rarely	1218	Sales	9	4	Life Sciences	1	27	3	Male	82	2	1
24	34	No	Travel_Rarely	419	Research & Development	7	4	Life Sciences	1	28	1	Female	53	3	3
25	21	No	Travel_Rarely	391	Research & Development	15	2	Life Sciences	1	30	3	Male	96	3	1
26	34	Yes	Travel_Rarely	699	Research & Development	6	1	Medical	1	31	2	Male	83	3	1
27	53	No	Travel_Rarely	1282	Research & Development	5	3	Other	1	32	3	Female	58	3	5
28	32	Yes	Travel_Frequently	1125	Research & Development	16	1	Life Sciences	1	33	2	Female	72	1	1
29	42	No	Travel_Rarely	691	Sales	8	4	Marketing	1	35	3	Male	48	3	2
30	44	No	Travel_Rarely	477	Research & Development	7	4	Medical	1	36	1	Female	42	2	3
31	46	No	Travel_Rarely	705	Sales	2	4	Marketing	1	38	2	Female	83	3	5

The Data – What is the Problem/Goal?

1. How well can you predict attrition based on other characteristics (e.g., age)?

Build a model, e.g., if $age \geq 65 \Rightarrow attrition=yes$

2. What drives attrition?

For example, age

3. What other general insights can you obtain from the data?

E.g., what distinguishes high performers?

	A	B	C	D	E	F	G
1	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromOffice	Education
2	41	Yes	Travel_Rarely	1102	Sales		1
3	49	No	Travel_Frequently	279	Research & Development		8
4	37	Yes	Travel_Rarely	1373	Research & Development		2
5	33	No	Travel_Frequently	1392	Research & Development		3
6	27	No	Travel_Rarely	591	Research & Development		2
7	32	No	Travel_Frequently	1005	Research & Development		2
8	59	No	Travel_Rarely	1324	Research & Development		3
9	30	No	Travel_Rarely	1358	Research & Development		24
10	38	No	Travel_Frequently	216	Research & Development		23
11	36	No	Travel_Rarely	1299	Research & Development		27
12	35	No	Travel_Rarely	809	Research & Development		16
13	29	No	Travel_Rarely	153	Research & Development		15
14	31	No	Travel_Rarely	670	Research & Development		26
15	34	No	Travel_Rarely	1346	Research & Development		19
16	28	Yes	Travel_Rarely	103	Research & Development		24
17	29	No	Travel_Rarely	1389	Research & Development		21
18	32	No	Travel_Rarely	334	Research & Development		5
19	22	No	Non-Travel	1123	Research & Development		16
20	53	No	Travel_Rarely	1219	Sales		2
21	38	No	Travel_Rarely	371	Research & Development		2
22	24	No	Non-Travel	673	Research & Development		11
23	36	Yes	Travel_Rarely	1218	Sales		9
24	34	No	Travel_Rarely	419	Research & Development		7
25	21	No	Travel_Rarely	391	Research & Development		15
26	34	Yes	Travel_Rarely	699	Research & Development		6
27	53	No	Travel_Rarely	1282	Research & Development		5
28	32	Yes	Travel_Frequently	1125	Research & Development		16
29	41	No	Travel_Rarely	691	Sales		8

Tableau – Downloading

The screenshot shows the Tableau Public website interface. At the top, there is a navigation bar with the Tableau Public logo and links for GALLERY, AUTHORS, BLOG, RESOURCES, ACTIVITY, and a SIGN IN button. The main content area features the heading "You'll be exploring in minutes" followed by a sub-heading: "Create and share interactive charts and graphs, stunning maps, live dashboards and fun applications in minutes, then publish anywhere on the web. Anyone can do it, it's that easy—and it's free." Below this is a form with an email input field containing "mvoortman@pointpark.edu" and a green "Download the App" button. A red box highlights the email field and the button. Underneath the button, it says "2019.1 Available for Windows and Mac (OS X) | System Requirements". The lower portion of the page shows a tablet displaying a Tableau dashboard titled "Taxi pickups" with a heatmap visualization of a city street grid. Below the tablet, the text "Feature Highlights" is centered. At the bottom, there are two feature highlights: "Heatmaps" with a small icon and the text "With one click, turn millions of marks into a", and "Transparent Sheets" with a small icon and the text "Unleash your creativity to build presentation-".

tableau+public

GALLERY AUTHORS BLOG RESOURCES ACTIVITY SIGN IN

You'll be exploring in minutes

Create and share interactive charts and graphs, stunning maps, live dashboards and fun applications in minutes, then publish anywhere on the web. Anyone can do it, it's that easy—and it's free.

mvoortman@pointpark.edu Download the App

2019.1 Available for Windows and Mac (OS X) | System Requirements

Feature Highlights

Heatmaps
With one click, turn millions of marks into a

Transparent Sheets
Unleash your creativity to build presentation-

Download from <https://public.tableau.com/en-us/s/download>

Tableau – Installing

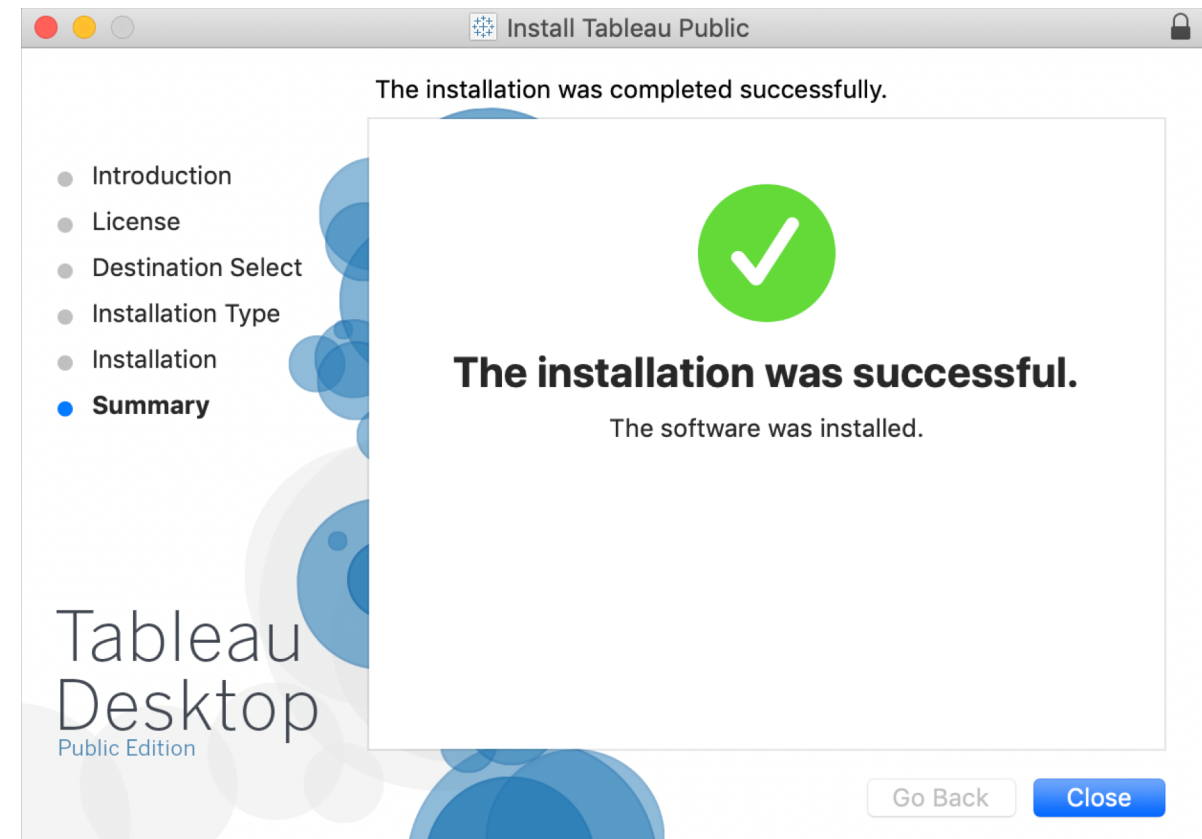
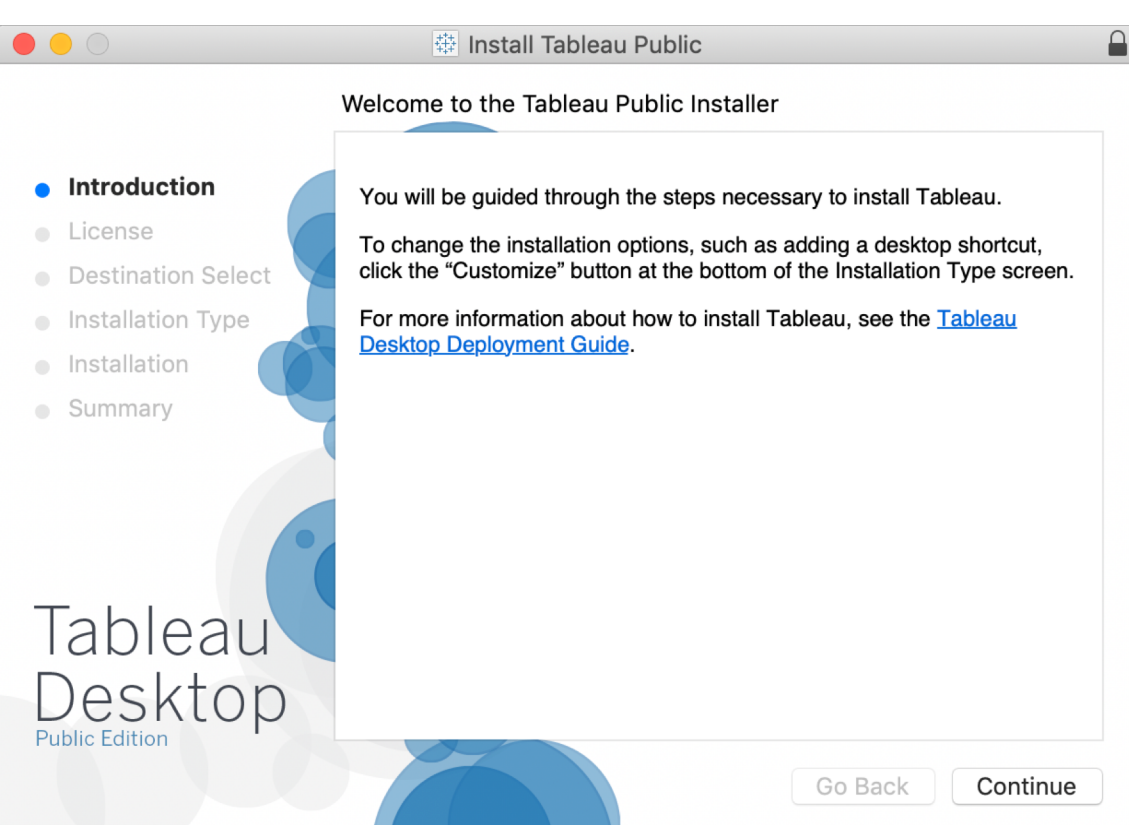
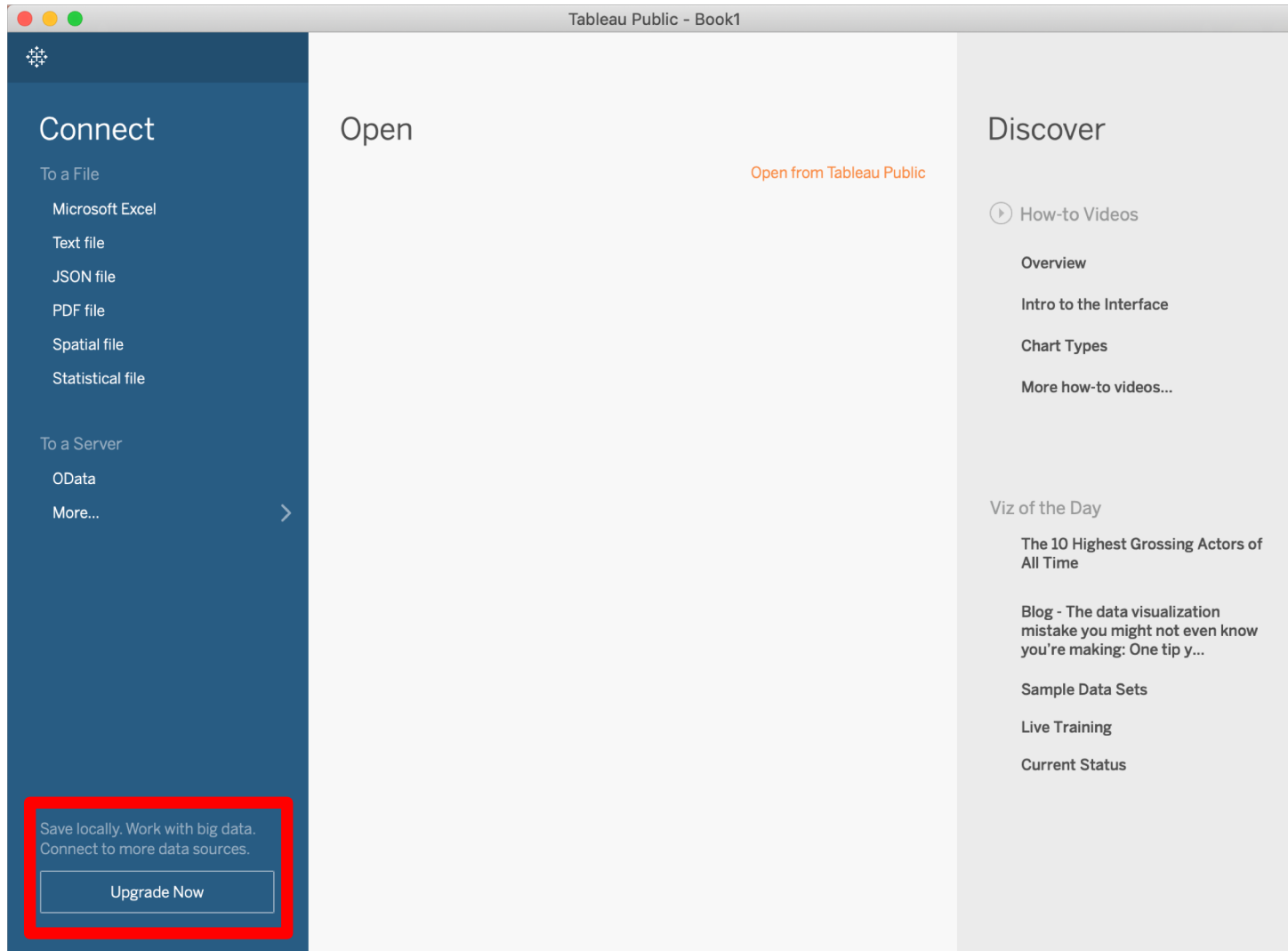
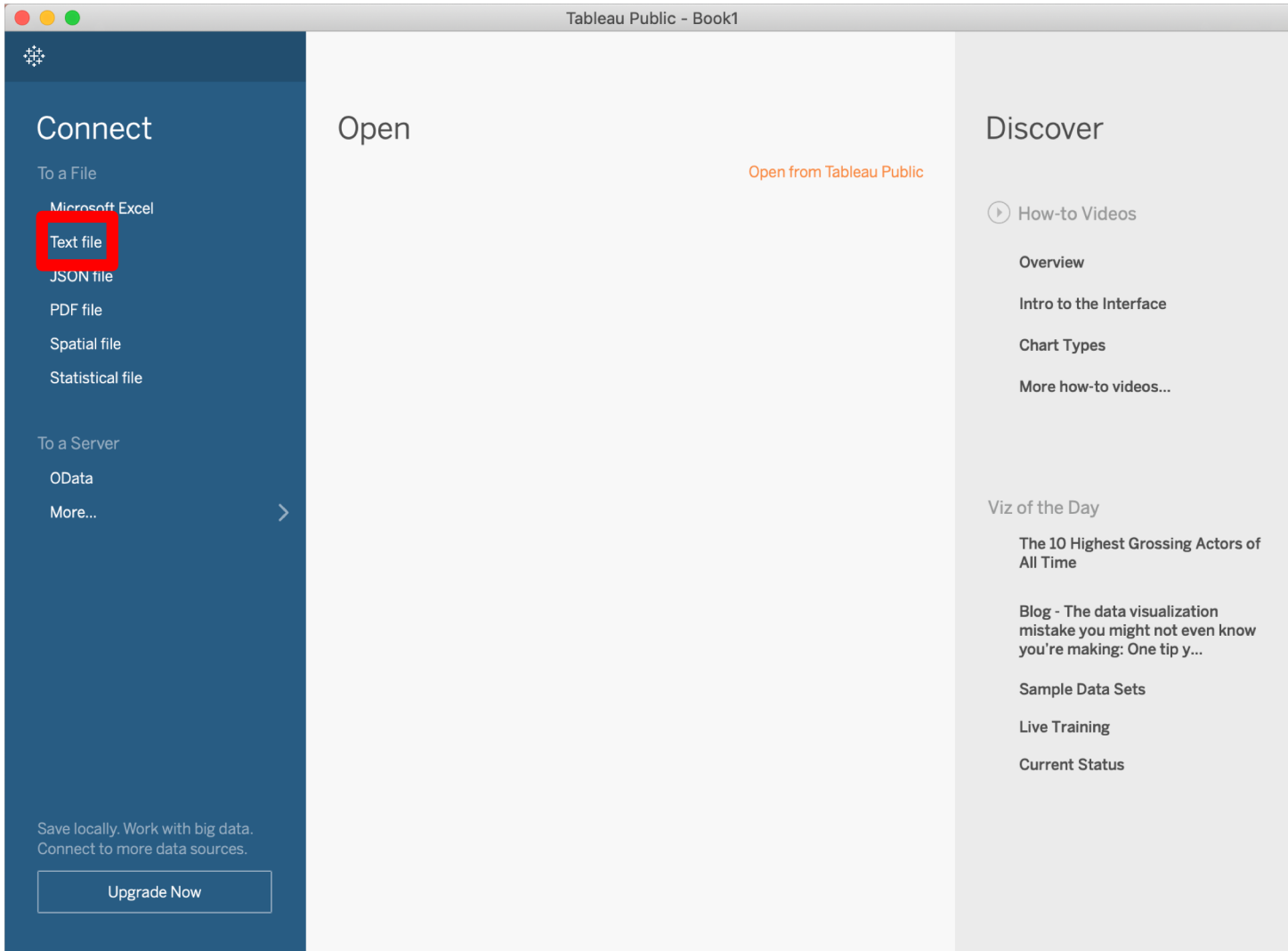


Tableau – Starting



Note: the public version has several limitations, such as not being able to save your work locally

Tableau – Loading the Data



Click 'Text file'

Then select
hr-employee-attribution.csv
and click open

Tableau – The Data Overview

The screenshot shows the Tableau Public interface with a data source overview for 'hr-employee-attrition.csv'. The interface includes a sidebar with 'Connections' and 'Files' sections, and a main area displaying the data source name and a preview table. The table is highlighted with a red border.

Tableau Public - Book1

hr-employee-attrition

Filters: 0 | Add

hr-employee-attrition.csv

Sort fields: Data source order | Show aliases | Show hidden fields | 1,000 rows

#	Age	Attrition	Business Travel	Daily Rate	Department	Distance From Ho...	Edu
41	41	Yes	Travel_Rarely	1,102	Sales	1	
49	49	No	Travel_Frequently	279	Research & Develop...	8	
37	37	Yes	Travel_Rarely	1,373	Research & Develop...	2	
33	33	No	Travel_Frequently	1,392	Research & Develop...	3	
27	27	No	Travel_Rarely	591	Research & Develop...	2	
32	32	No	Travel_Frequently	1,005	Research & Develop...	2	
59	59	No	Travel_Rarely	1,324	Research & Develop...	3	
30	30	No	Travel_Rarely	1,358	Research & Develop...	24	
38	38	No	Travel_Frequently	216	Research & Develop...	23	
36	36	No	Travel_Rarely	1,299	Research & Develop...	27	

Here is our data!

Tableau – The Variable Overview

The screenshot shows the Tableau Public interface for a workbook named "Tableau Public - Book1". The main view displays the data source "hr-employee-attrition.csv". On the left sidebar, there are sections for "Connections" (showing "hr-employee-attrition Text file"), "Files" (with a checkbox for "Use Data Interpreter" and a note about cleaning text file workbooks), and "New Union". Below the sidebar, there is a "Sort fields" dropdown menu set to "Data source order" and a "Show hidden fields" checkbox. A red box highlights the "Sort fields" icon and the table below it.

Field Name	Table	Remote Field Name
# Age	hr-employee-attrition.csv	Age
Abc Attrition	hr-employee-attrition.csv	Attrition
Abc Business Travel	hr-employee-attrition.csv	BusinessTravel
# Daily Rate	hr-employee-attrition.csv	DailyRate
Abc Department	hr-employee-attrition.csv	Department
# Distance From Home	hr-employee-attrition.csv	DistanceFromHome
# Education	hr-employee-attrition.csv	Education
Abc Education Field	hr-employee-attrition.csv	EducationField
# Employee Count	hr-employee-attrition.csv	EmployeeCount
# Employee Number	hr-employee-attrition.csv	EmployeeNumber
# Environment Satisfaction	hr-employee-attrition.csv	EnvironmentSatisfaction

Here are our variables!

Tableau – Going to the Worksheet

Tableau Public - Book1

hr-employee-attrition

Filters
0 | Add

hr-employee-attrition.csv

Connections
Add
hr-employee-attrition
Text file

Files
Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.
hr-employee-attrition.csv

New Union

Sort fields: Data source order

Show hidden fields

Field Name	Table	Remote Field Name
# Age	hr-employee-attrition.csv	Age
Abc Attrition	hr-employee-attrition.csv	Attrition
Abc Business Travel	hr-employee-attrition.csv	BusinessTravel
# Daily Rate	hr-employee-attrition.csv	DailyRate
Abc Department	hr-employee-attrition.csv	Department
# Distance From Home	hr-employee-attrition.csv	DistanceFromHome
# Education	hr-employee-attrition.csv	Education
Abc Education Field	hr-employee-attrition.csv	EducationField
# Employee Count	hr-employee-attrition.csv	EmployeeCount
# Employee Number	hr-employee-attrition.csv	EmployeeNumber
# Environment Satisfaction	hr-employee-attrition.csv	EnvironmentSatisfaction

Go to Worksheet

Sheet 1

**That is where we create
all figures and plots**

Tableau – The Worksheet

The screenshot displays the Tableau Public interface for a workbook titled "Tableau Public - Book1". The interface is divided into several key sections:

- Top Bar:** Contains navigation icons (back, forward, home, refresh), a "Standard" view dropdown, and a "Show Me" button.
- Data Pane (Left):** Shows the data source "hr-employee-attrition". It is categorized into:
 - Dimensions:** Attrition, Business Travel, Department, Education Field, Employee Number, Gender, Job Role, Marital Status, Over Time, Over18, Training Times Last Year, and Measure Names.
 - Measures:** Age, Daily Rate, Distance From Home, Education, Employee Count, Environment Satisfaction, Hourly Rate, Job Involvement, Job Level, Job Satisfaction, Monthly Income, Monthly Rate, and Num Companies Worked.
- Columns and Rows Shelves:** Both are currently empty, with the text "Drop field here" indicating where to place dimensions or measures.
- Marks Card (Left):** Shows the mark type set to "Automatic". Below it are options for "Color", "Size", and "Text" (all currently empty), and "Detail" and "Tooltip" (both containing icons).
- Visualization Gallery (Right):** A grid of various chart types including bar charts, line graphs, pie charts, maps, and treemaps. Below the gallery is the text: "Select or drag data. Use the Shift or Cmd key to select multiple fields".
- Bottom Bar:** Shows the "Data Source" and "Sheet 1" tabs, along with navigation icons for the worksheet.

This looks complicated!

Tableau – Our First Visualization (Bar Chart)

The screenshot shows the Tableau Public interface with the 'hr-employee-attrition' data source loaded. The 'Attrition' field is highlighted in the Dimensions pane and is being dragged to the Columns shelf, as indicated by a red box around the 'Attrition' dropdown in the Columns shelf. The Marks shelf is currently set to 'Automatic'. The main view shows a table with columns for 'Attrition' and 'Employee Number'.

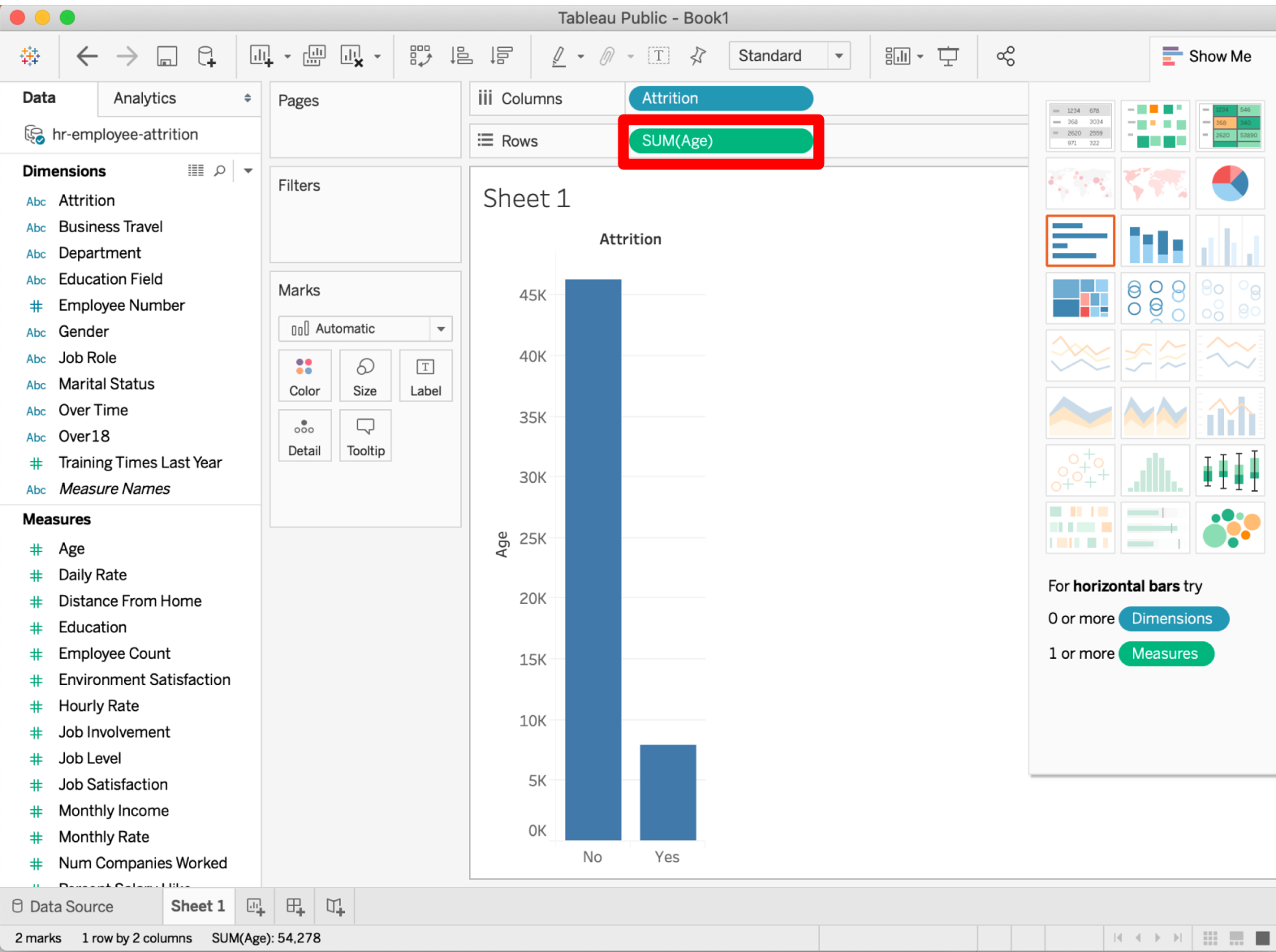
Attrition	
No	Yes
Abc	Abc

Age vs. Attrition

Step 1

Drag Attrition to Columns

Tableau – Our First Visualization (Bar Chart)



Age vs. Attrition

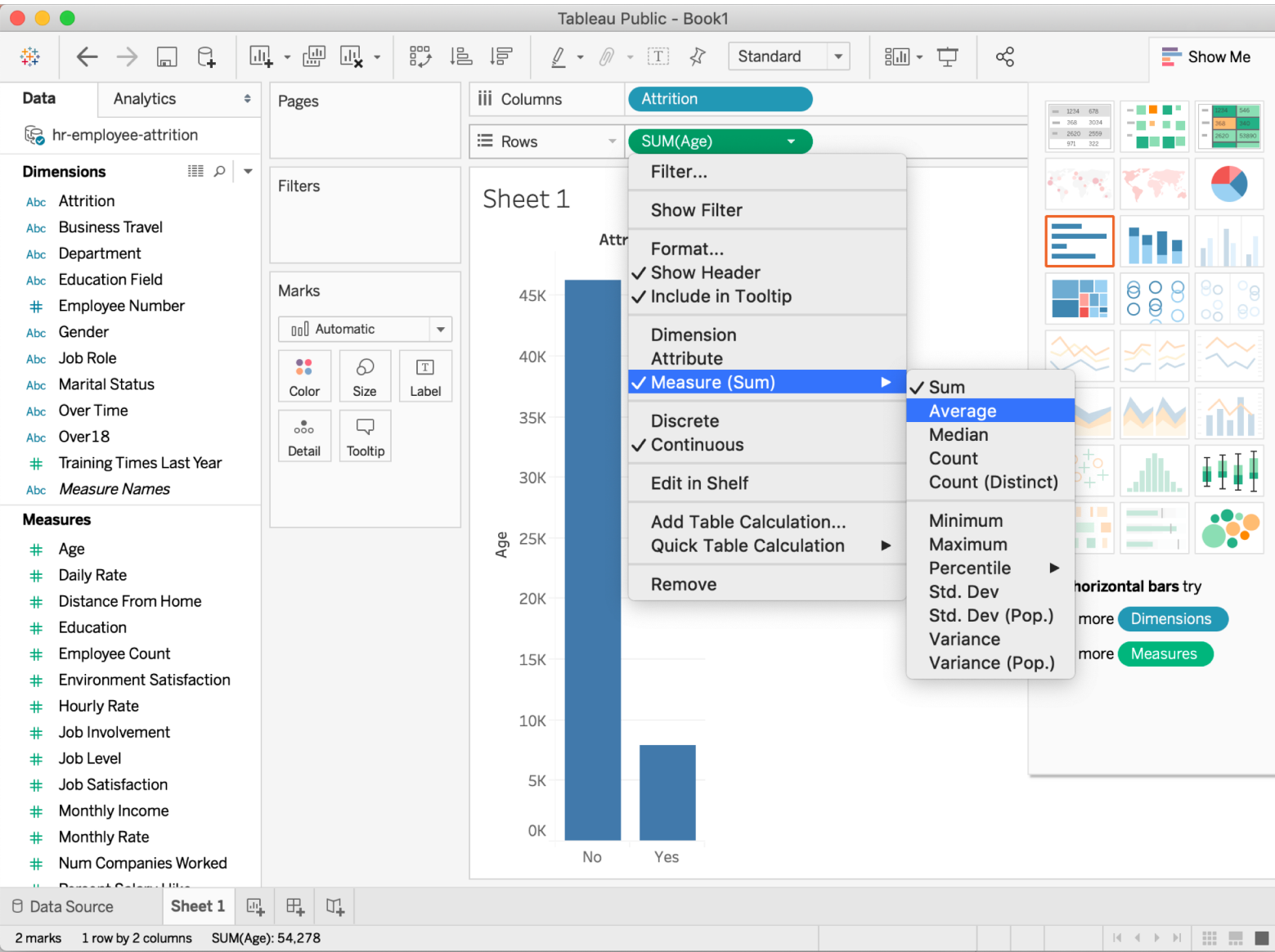
Step 2

Drag Age to Rows

Problem!

Age is summed!

Tableau – Our First Visualization (Bar Chart)

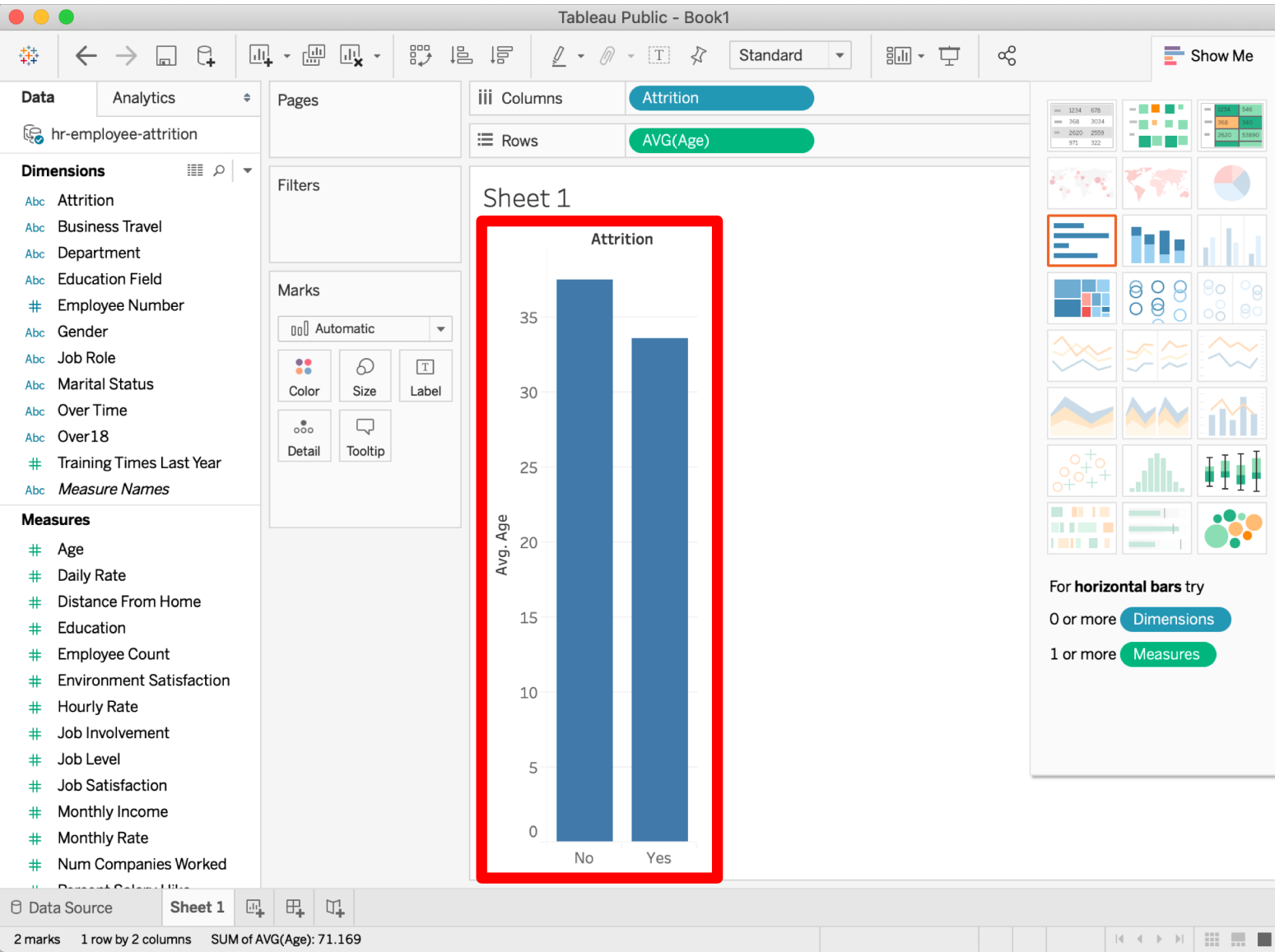


Age vs. Attrition

Step 3

Change sum to average

Tableau – Our First Visualization (Bar Chart)



Age vs. Attrition

Step 4

The end result!

Employees that leave are younger than employees that stay (on average)

Tableau – Clearing the Visualization

The screenshot shows the Tableau Public interface with a bar chart titled 'Attrition' on 'Sheet 1'. The chart displays 'Avg. Age' on the y-axis (ranging from 0 to 35) and 'Attrition' on the x-axis (with categories 'No' and 'Yes'). The 'Attrition' field is on the Columns shelf and 'AVG(Age)' is on the Rows shelf. The 'Clear Sheet' button in the top toolbar is highlighted with a red box. The interface also shows the Data pane on the left with dimensions like Attrition, Business Travel, Department, etc., and measures like Age, Daily Rate, etc. The status bar at the bottom indicates '2 marks 1 row by 2 columns SUM of AVG(Age): 71.169'.

Attrition	Avg. Age
No	~37
Yes	~34

**Click the Clear Sheet
item to restart**

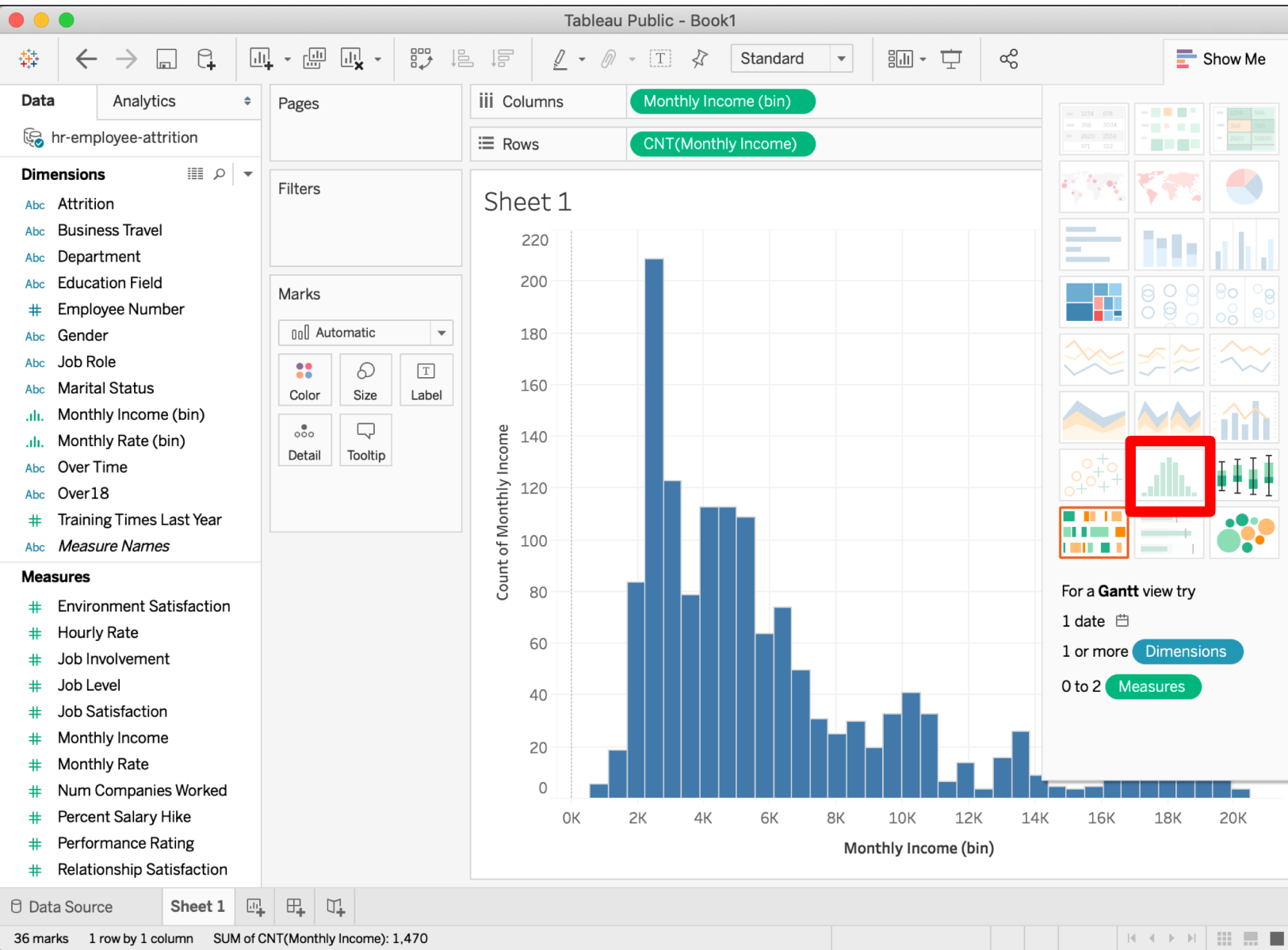
Tableau – Creating a Histogram

The screenshot shows the Tableau Public interface with the following components:

- Top Bar:** Includes navigation icons, a 'Standard' view dropdown, and a 'Show Me' button.
- Data Pane:** Shows the data source 'hr-employee-attrition'.
- Dimensions:** A list of fields including Attrition, Business Travel, Department, Education Field, Employee Number, Gender, Job Role, Marital Status, Monthly Income (bin), Monthly Rate (bin), Over Time, Over18, Training Times Last Year, and Measure Names.
- Measures:** A list of fields including Environment Satisfaction, Hourly Rate, Job Involvement, Job Level, Job Satisfaction, Monthly Income, Monthly Rate, Num Companies Worked, Percent Salary Hike, Performance Rating, and Relationship Satisfaction.
- Columns and Rows:** Both are currently empty.
- Marks Card:** The mark type is set to 'Automatic'. A green pill labeled 'SUM(Monthly I..)' is visible in the Marks card.
- Visualization Area:** The main workspace shows 'Sheet 1' with a single data point of 9,559,309.
- Show Me Panel:** A dropdown menu is open, showing various chart types. A histogram icon is highlighted with an orange box. Below the icons, there are instructions: 'For a Gantt view try 1 date', '1 or more Dimensions', and '0 to 2 Measures'.
- Bottom Bar:** Shows 'Data Source', 'Sheet 1', and a status bar indicating '1 mark 1 row by 1 column SUM(Monthly Income): 9,559,309'.

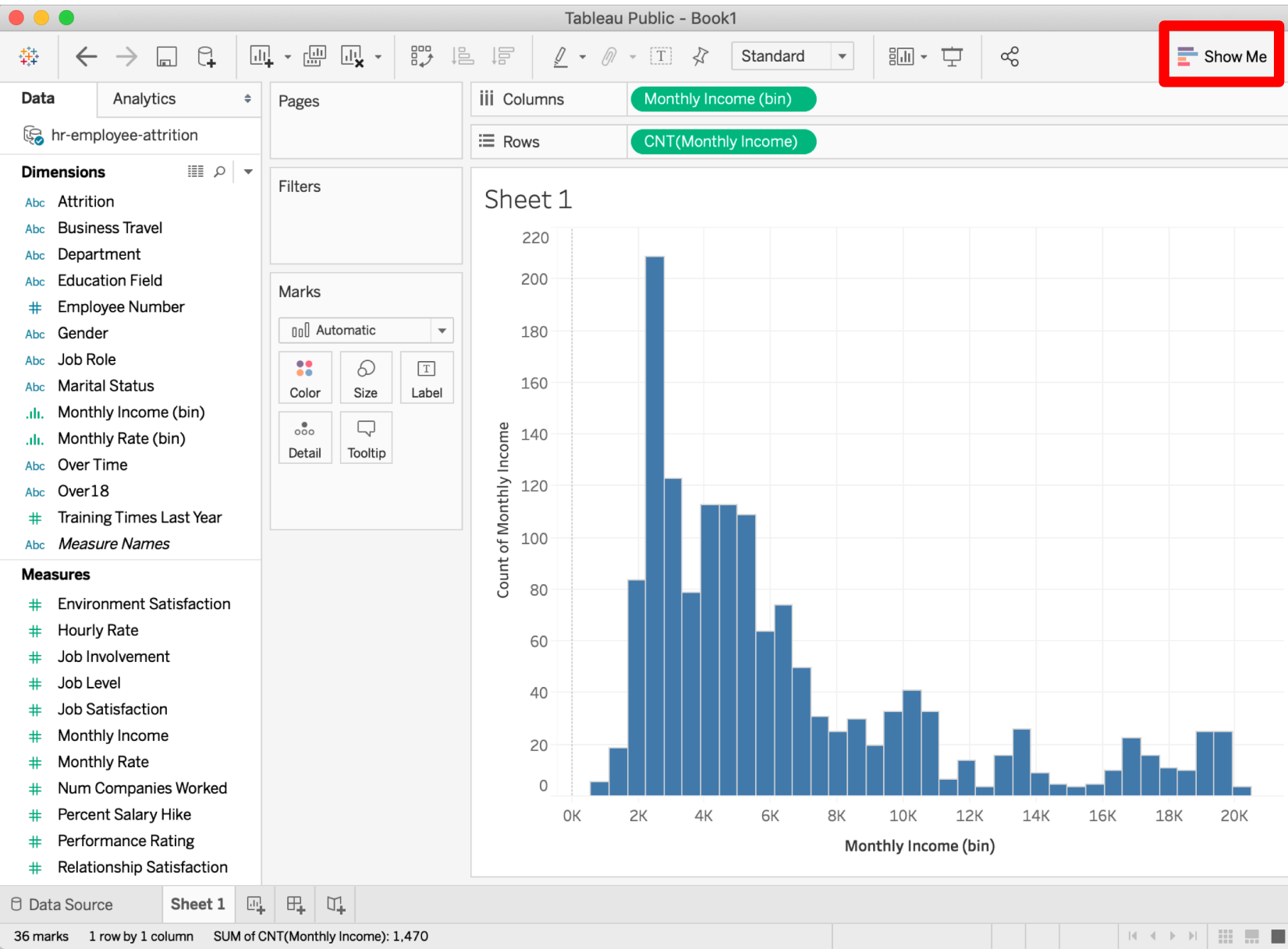
**Drop Monthly
Income in the
middle of the
worksheet**

Tableau – Creating a Histogram



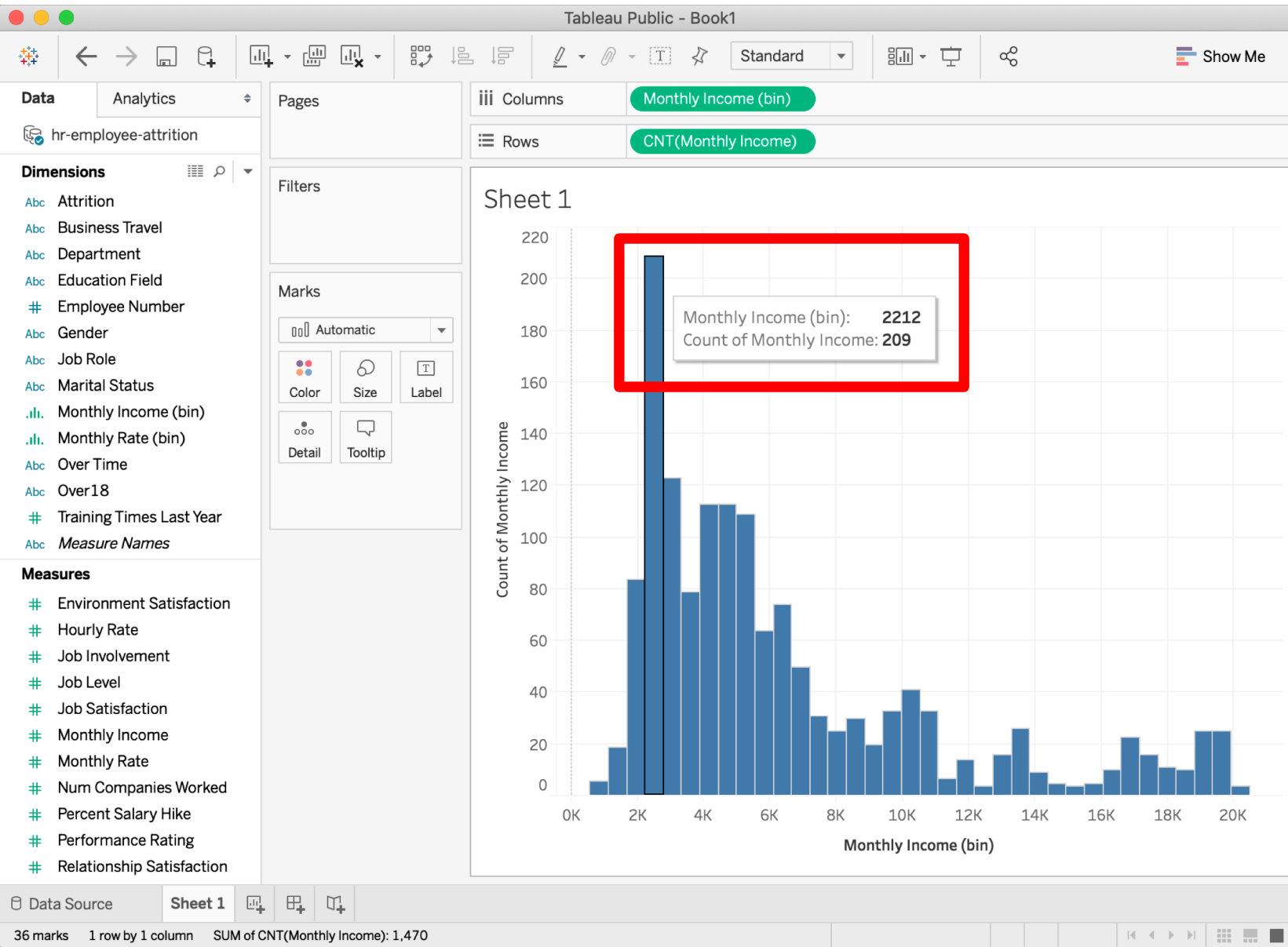
**Click on the
histogram icon**

Tableau – Creating a Histogram



Hide the menu by clicking on Show Me

Tableau – Creating a Histogram



**Hovering over
one of the bars**

Tableau – Exporting as PowerPoint

The screenshot shows the Tableau Public interface. The 'File' menu is open, and the 'Export As PowerPoint...' option is highlighted with a red box. The main view displays a histogram of 'Monthly Income (bin)' with a count of 1,470. The histogram shows a distribution of monthly income bins, with the highest frequency occurring between 2K and 4K. The x-axis is labeled 'Monthly Income (bin)' and ranges from 0K to 20K. The y-axis is labeled 'Count of Monthly Income' and ranges from 0 to 140. The Tableau Public interface includes a menu bar (File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Window, Help), a toolbar, and a sidebar with Dimensions and Measures.

Monthly Income (bin)	Count of Monthly Income
0K - 1K	5
1K - 2K	20
2K - 3K	85
3K - 4K	125
4K - 5K	80
5K - 6K	115
6K - 7K	110
7K - 8K	65
8K - 9K	50
9K - 10K	30
10K - 11K	25
11K - 12K	20
12K - 13K	15
13K - 14K	10
14K - 15K	25
15K - 16K	10
16K - 17K	5
17K - 18K	20
18K - 19K	15
19K - 20K	25
20K - 21K	25

**Export directly as
PowerPoint**

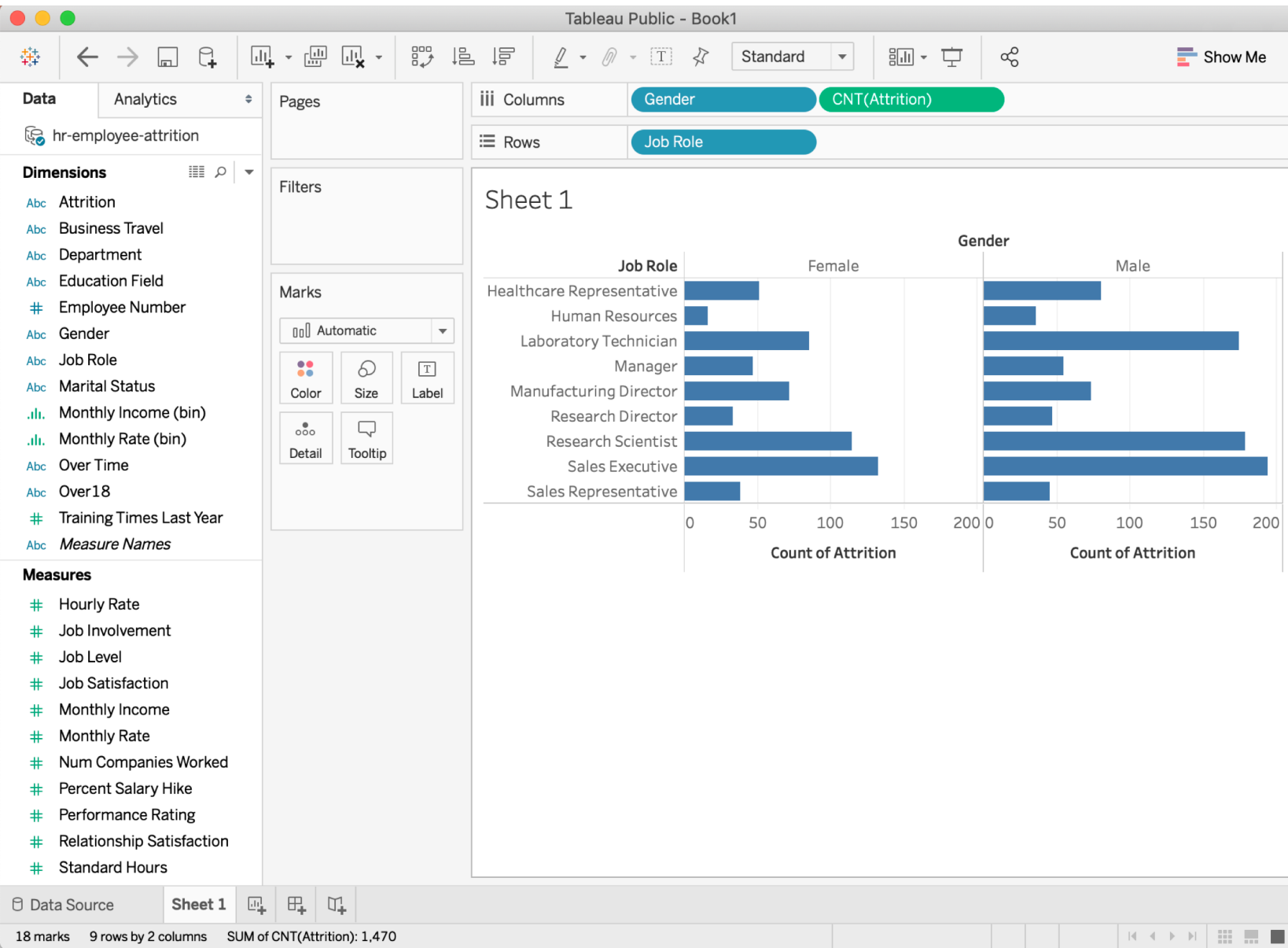
Tableau – Multiple Dimensions

The screenshot shows the Tableau Public interface with a data table. The table has columns for Job Role, Gender, and Attrition. A context menu is open over the Attrition field, showing options like Filter, Sort, and Measure. The Measure option is selected, and a sub-menu is open showing options like Minimum, Maximum, Count, and Count (Distinct). The Count option is highlighted.

Job Role	Gender	Attrition
Healthcare Representative	Female	No
Human Resources	Male	Yes
Laboratory Technician	Female	No
Manager	Male	Yes
Manufacturing Director	Female	No
Research Director	Male	Yes
Research Scientist	Female	No
Sales Executive	Male	Yes
Sales Representative	Female	No

**Gender vs. Job Role
vs. Attrition**

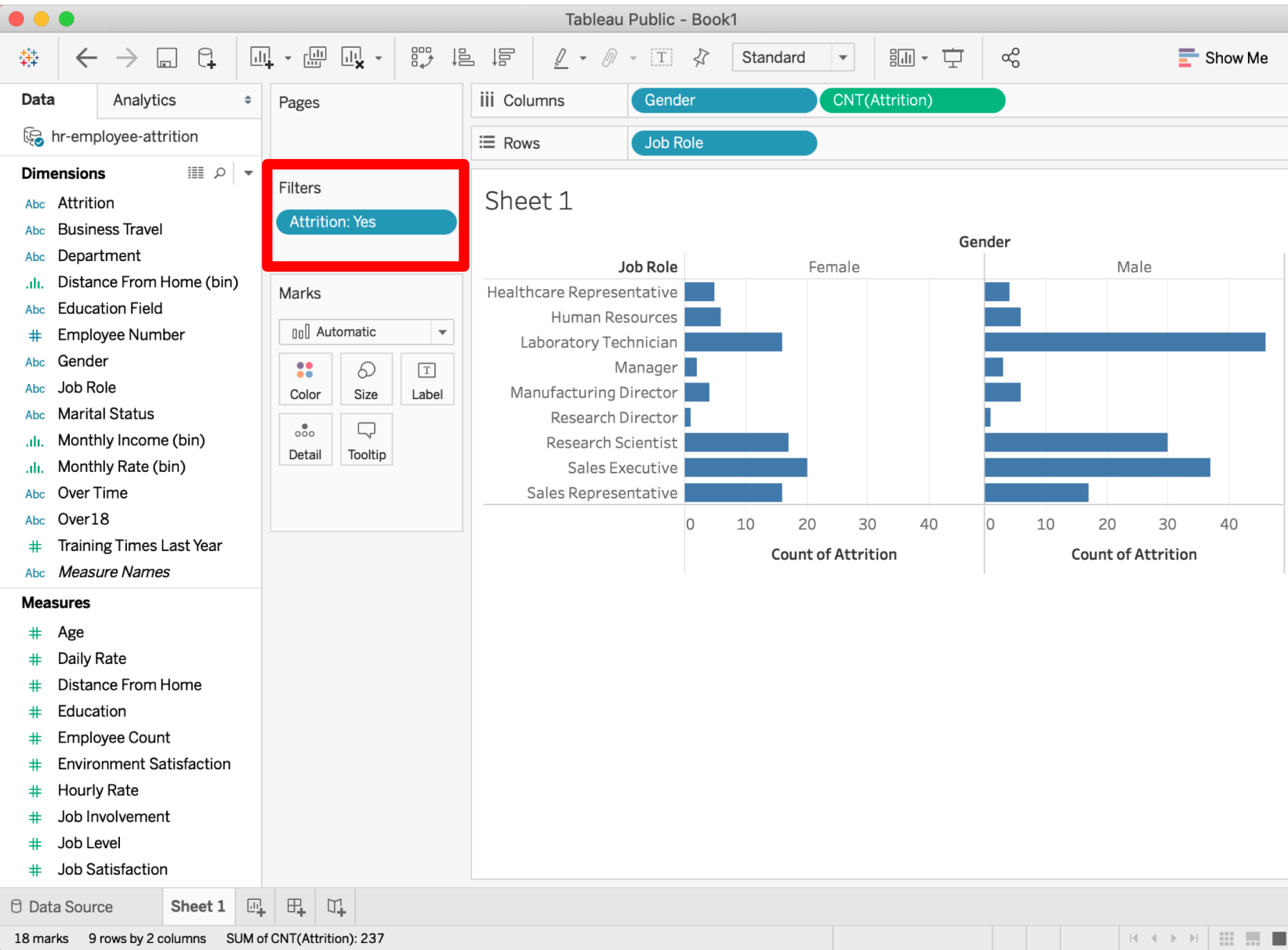
Tableau – Multiple Dimensions



**Gender vs. Job Role
vs. Attrition**

What do you think?

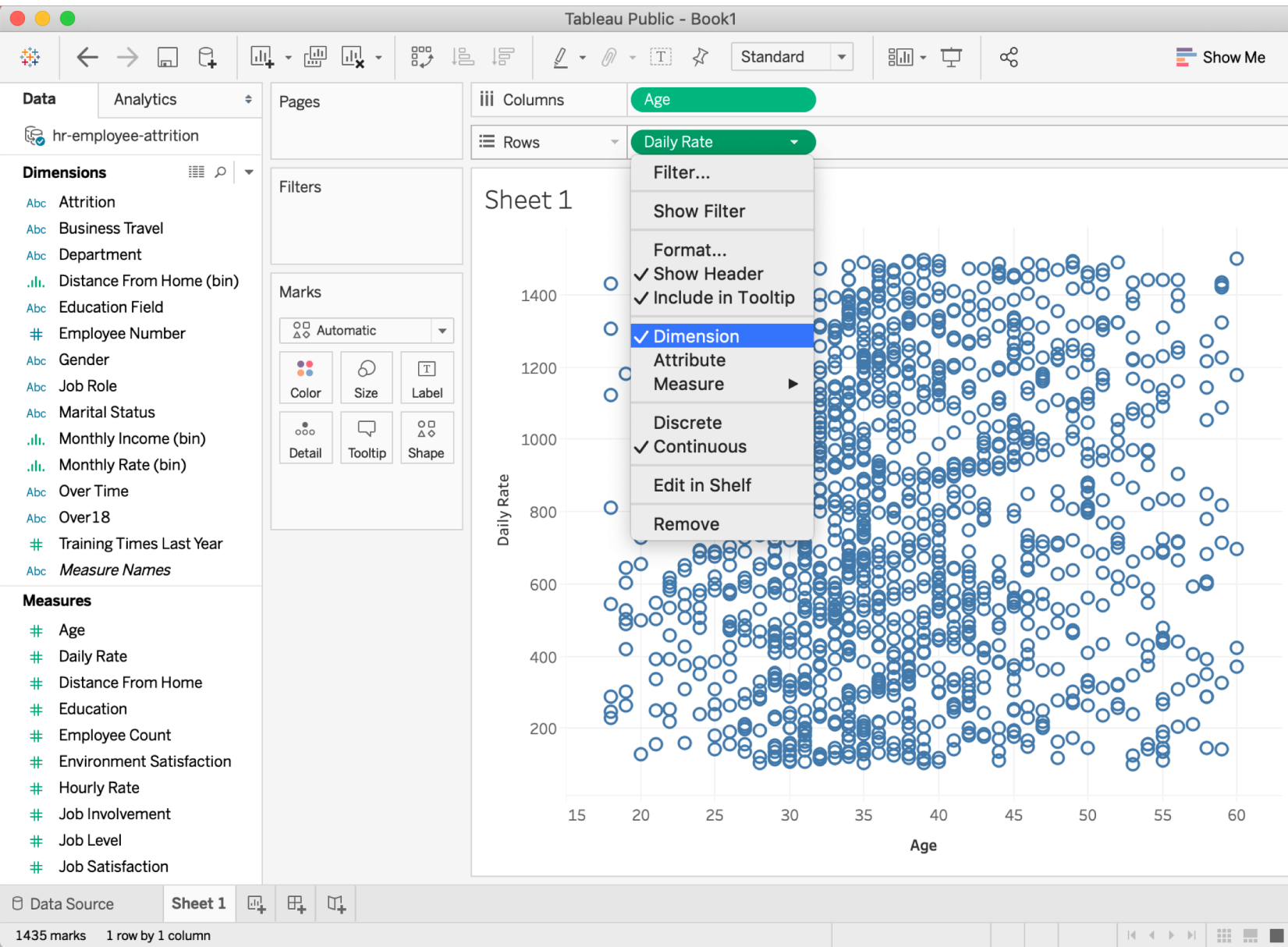
Tableau – Multiple Dimensions



**Gender vs. Job Role
vs. Attrition**

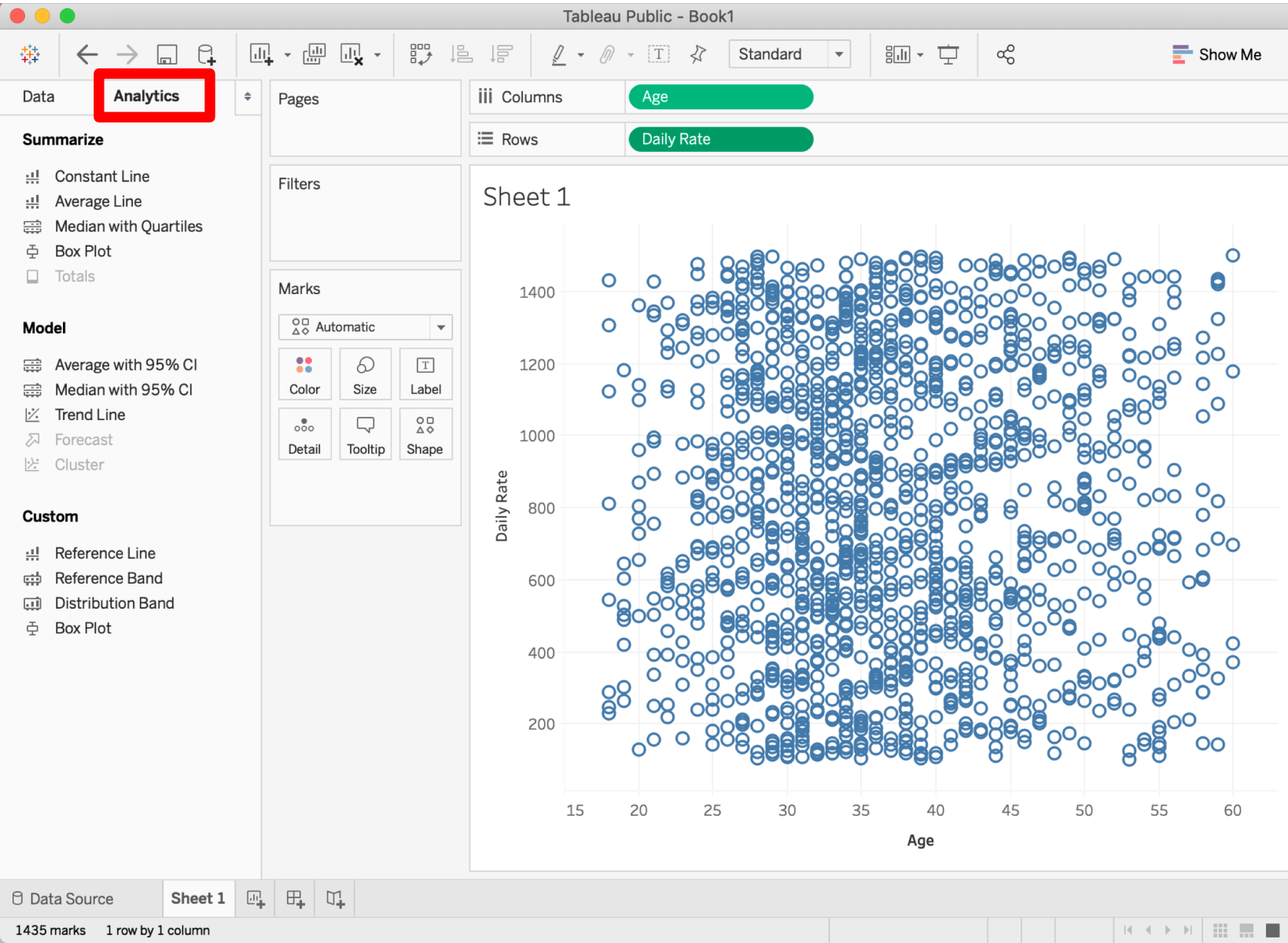
Filtering!

Tableau – A Scatter Plot



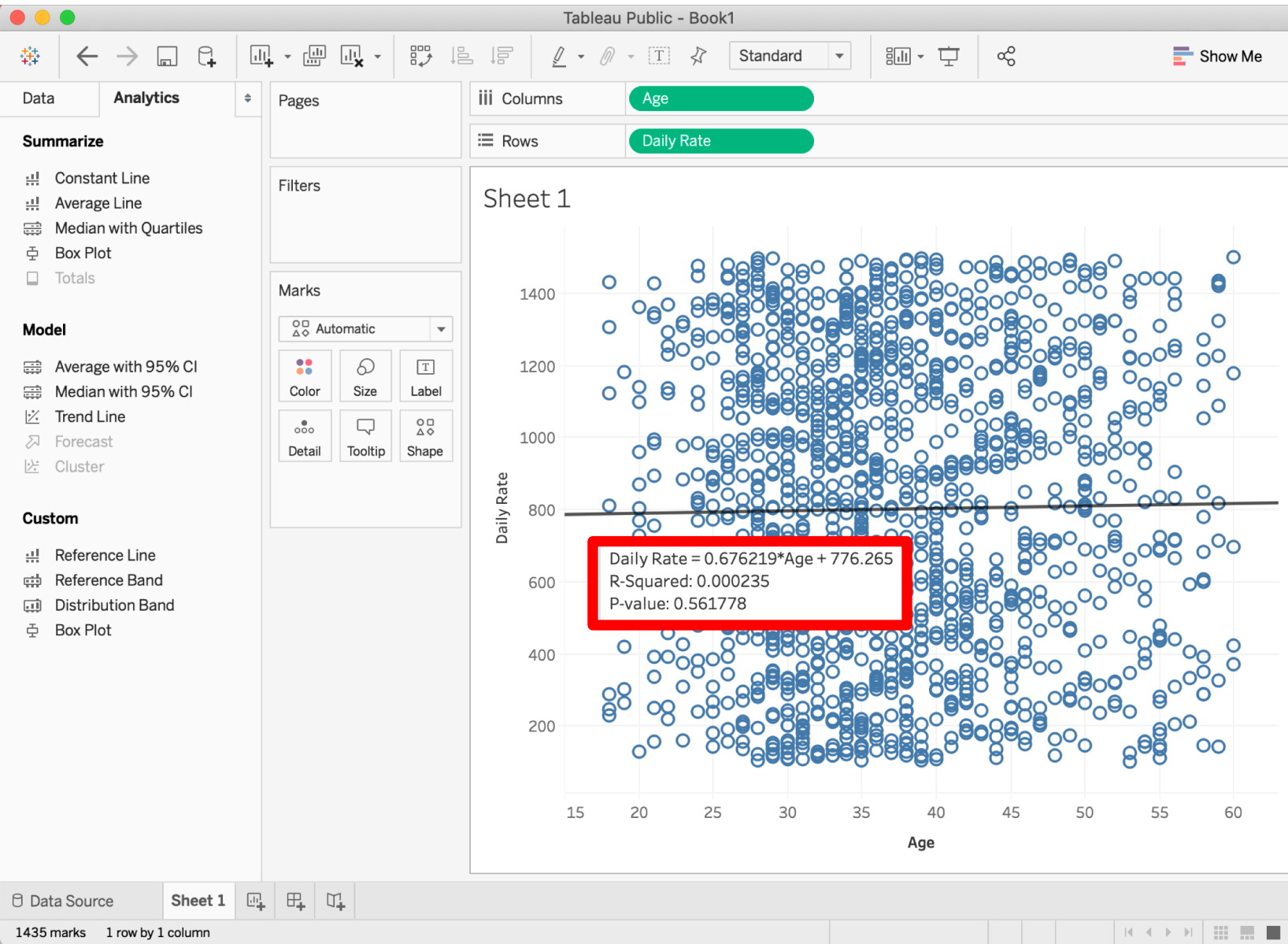
**A scatter plot of
Age vs. Daily Rate**

Tableau – A Scatter Plot



Go to the Analytics tab

Tableau – A Scatter Plot



**Click on Trend Line,
drag to the right,
and drop on Linear**

What do you think?

Next Steps

- Try it out!
- A **lot** of **trial** and **error** (try to be somewhat systematic)
- Use Google
- Ask questions