phuse·virtual

Data Science Theme
Hands-on Workshop

DSc

Data Science: Data Storytelling Hands-on Workshop

Sascha Ahrweiler, Karnika Dalal & Meenakshi Behl

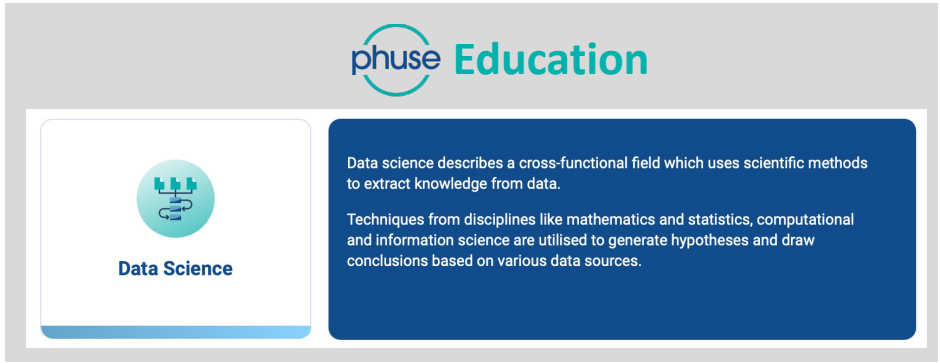
Bayer & Janssen Research & Development

Data Science Workflow

Extract
Transform
Load

Analyse
Statistical Modelling
Interpretation

Initiate Change



Why Data Storytelling?

Deliver valuable data insights in a way that stays in peoples mind to turn data into action

Image Source: https://www.forbes.com/sites/brentdykes/2016/07/13/data-storytelling-separating-fiction-from-facts-2/

Photo by Jakob Owens on Unsplash

Workshop Structure

- Welcome to Hollywood – the Art of Storytelling
- The **BIG IDEA**
- The 3 minute story
- The Audience
- From Data to Insights
- Turn insights into **visuals**
- **Storytelling**
- Storyboarding

Interactive

Chat

Breakout Room

Welcome to Hollywood
The Art of Storytelling

- People remember **stories** not data
- If you want that your data insights turn into action you have to deliver a story which stuck in peoples mind
- **Not a natural strength** of a Statistical Programmer
- If you listen to a good story you will **remember** it even after a long time
- Workshop should build some **fundamental** skills for effective storytelling
- Practice, practice, practice!

Photo by Matthew Osborn on Unsplash

Photo by Diego PH on Unsplash

Chat

The BIG IDEA

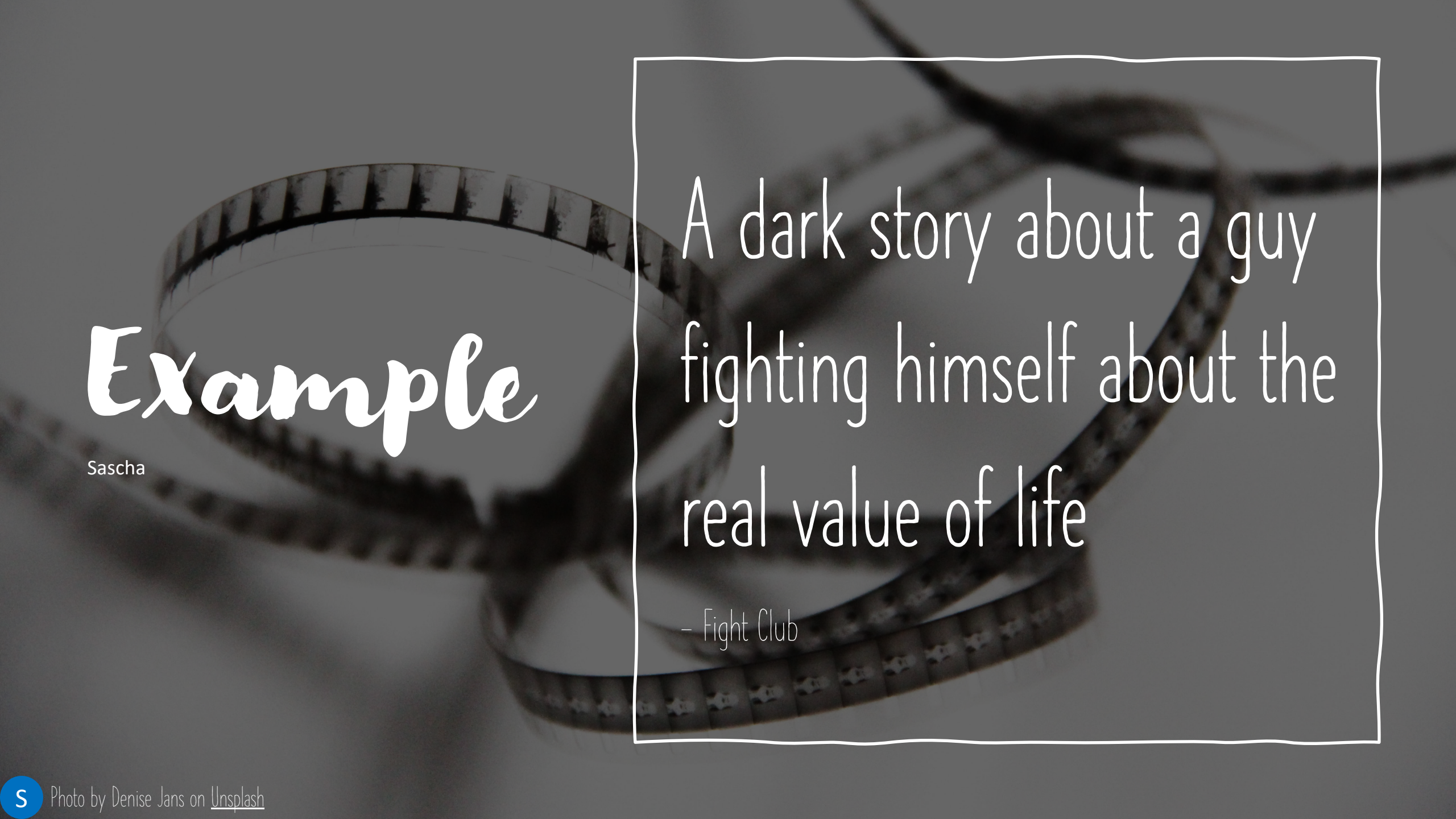
- Your audience should be able to **remember the big idea** behind your story
- The **BIG IDEA** is the core meaning of your story
- If it is too complicated people will not get it

Interactive:

What is your favorite movie?

Tell us the BIG IDEA behind it in **1** sentence!

Post it into the Teams chat



Example

Sascha

A dark story about a guy fighting himself about the real value of life

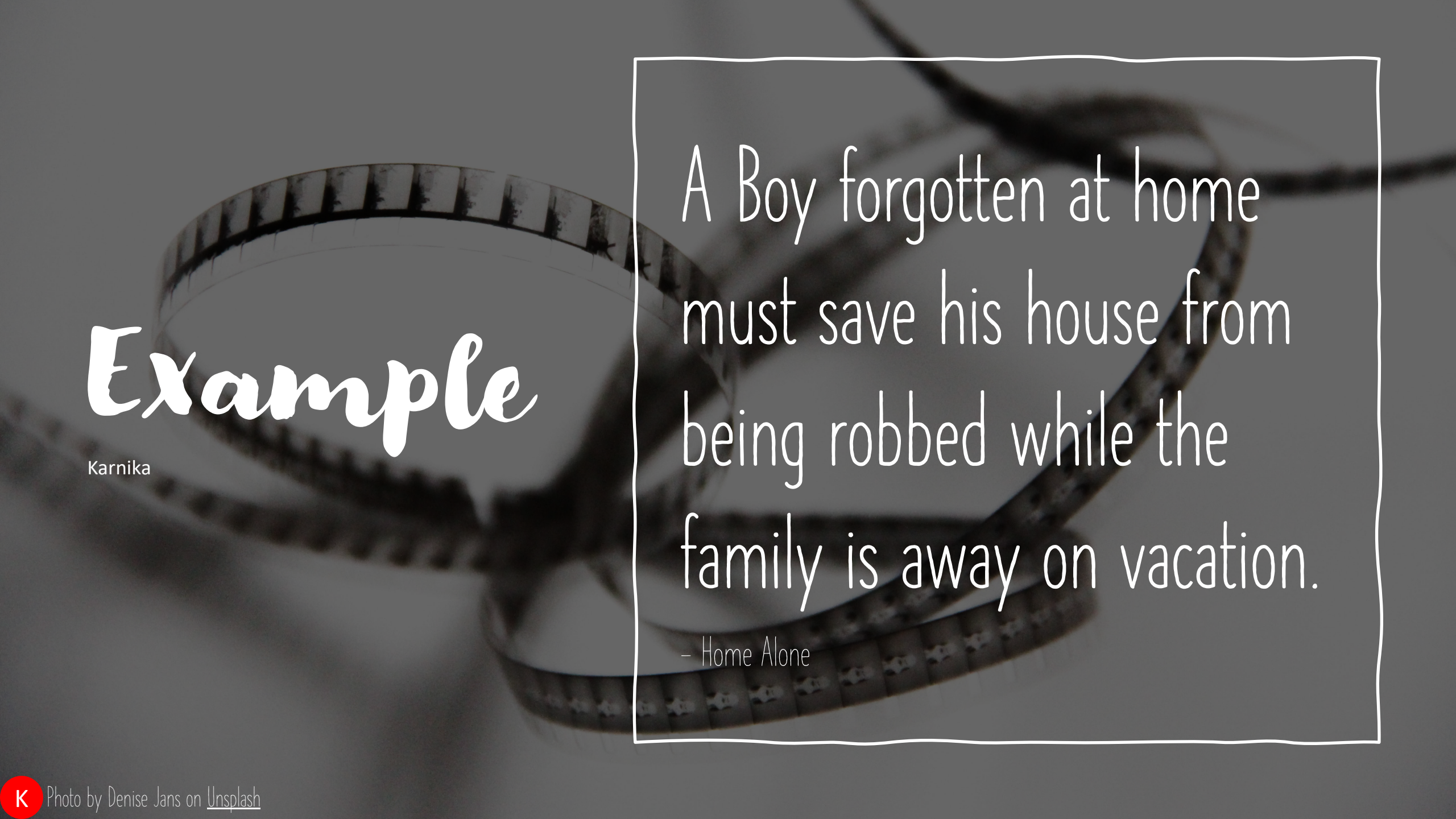
– Fight Club

Example

Karnika

A Boy forgotten at home must save his house from being robbed while the family is away on vacation.

– Home Alone

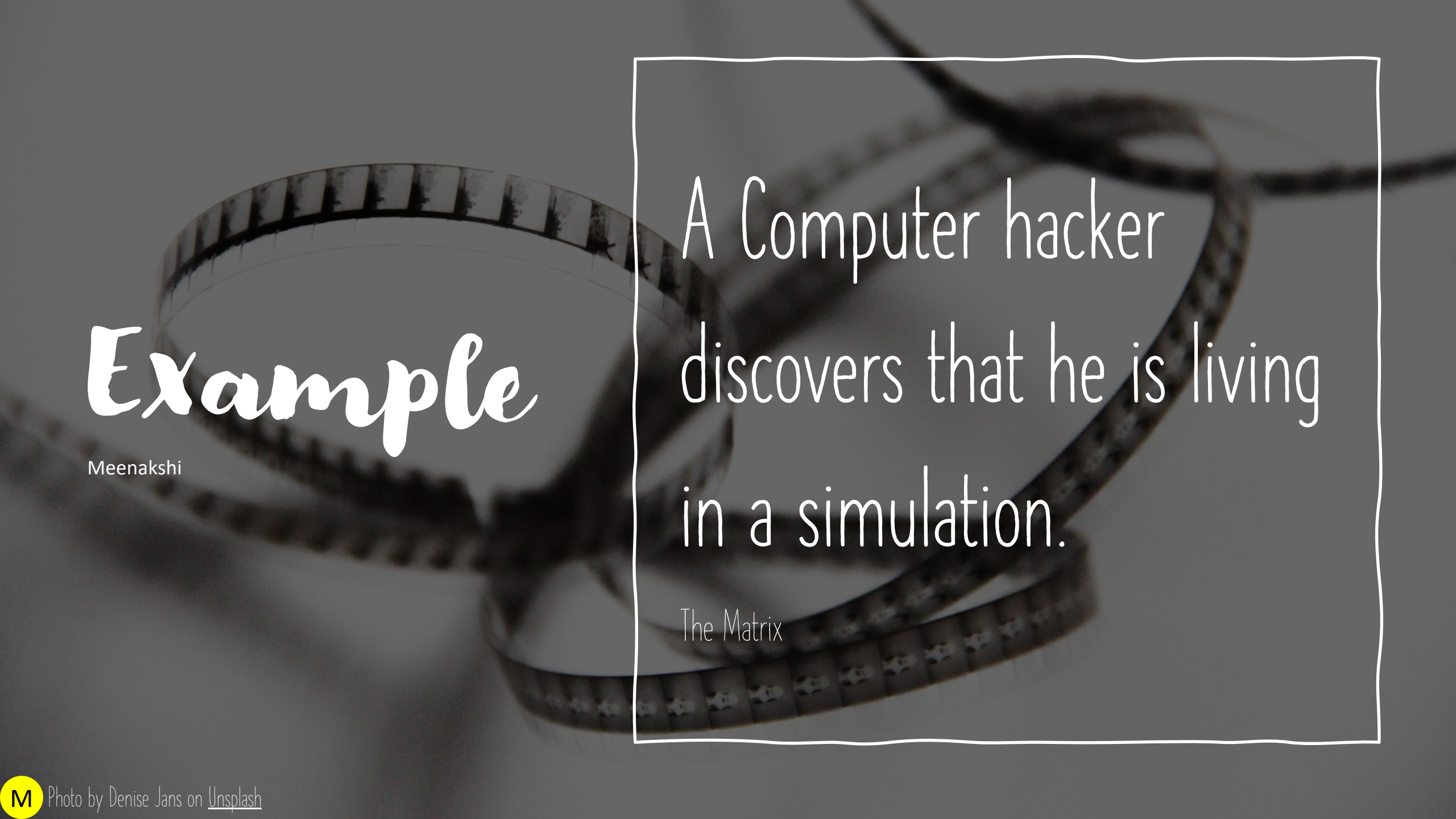


Example

Meenakshi

A Computer hacker discovers that he is living in a simulation.

The Matrix



Learning

- Getting to the real core of your message is **not easy**
- Thinking about the core idea will **help you to deliver** it better to your audience
- You want the audience to **keep your core idea in mind** and turn it into action

Photo by James Pond on Unsplash

Photo by Sung Jon Chi on Unsplash

Coming Soon

Breakout Room

The 3 minute story

- What if you would meet your boss in an elevator and you only would have 3 minutes to tell him what you are working on?
- If you are able to tell your story in 3 minutes it keeps you independent from slides
- It also helps to focus on the most important pieces of your story and reduce all the clutter
- Practice: We will assign you to Breakout rooms named after a movie
- Your tasks:

✓ Try to **talk** about a short narrative, which explains the story in about 3 minutes

✓ You will be pulled back in 10 minutes

Coming Soon

Breakout Room

Breakout Room Task

- Your Breakout Room has the name of a very **famous movie**
- In the next **10 minutes**, try to **tell the storyline** of this movie with the goal to tell this story **in 3 minutes**
- The structure of the story should **follow the narrative structure of the movie** like:
  - Main character introduction
  - What happens to this character? Any problems or external threats?
  - What does the character do to solve the situation?
  - Was the character successful at the end?

Can you build a really exciting story, which will fascinate the listener as the movie did?

K

Coming Soon

Breakout Room

Breakout to Movie Rooms

- Star Wars
- Lord of the Rings
- The Matrix
- E.T. – The Extraterrestrial
- Jurassic Park
- Iron Man
- The Lion King
- Toy Story
- Snow White and the seven Dwarfs
- The Beauty and the Beast
- Etc.

K

Learning

- Needs practice, practice, practice…
- You will need to learn how to articulate your story as **clearly and precisely**
- This will make it **easier to choose** the right content and **visuals** for your data story
- We will talk about the storyboarding later and will learn how to ensure that your audience will stay focused and follow your train of thoughts
- During this exercise, was there anyone who did not have the **visuals of the movie** in their mind?

Photo by Ricardo Annandale on Unsplash

Photo by Krist Luhaers on Unsplash

The Audience

- When telling your story it is essential to think about the **context** of your story
- Your story needs to **fit to your audience** and what you want to achieve
- When you want to present your data insights think about how to present it to your
  - Direct Reports
  - Peers
  - Senior Leadership
- The BIG IDEA and the 3-minute story will stay almost the same
- The details and the way you present will differ though

Learning

- **BIG IDEA** – the core idea of your story
- **3-minute story** – be precise and clear, be independent from any slides
- Think about your **audience**

Photo by Denise Jans on Unsplash

What's Next?

- From Data to Insights
- Turn insights into **visuals**
- **Storytelling**
- Storyboarding

DISCLAIMER

We will use demographic and baseline characteristic data from COVID19 vaccine studies. This is for demonstration purposes only. All vaccines have undergone a thorough review by global agencies with regards to the safety and efficacy of the vaccines. No concerns with regards to safety or efficacy were raised during the review.

We do not question the thorough agency review or the conclusions in any shape or form nor do we intend to blame any company to not adhere to high quality standards in the planning and conduct of clinical studies.

Photo by Clementina Toso on Unsplash

Background

- Starting November 2020, FDA and EMA approved several vaccines to fight the Covid-19 pandemic
- Despite a positive vote by the EMA, French president Macron publicly stated that "*the AstraZeneca Covid-19 vaccine is in-effective in the elderly population*"

A Data Story

- Now that you know the fundamentals of storytelling it is time to pick your main actor: **DATA**
- As a Clinical Data Scientist you know your domain inside out
- You spend weeks to extract, transform, load and analyse the data
- Is your audience interested in everything you did to get to a single number, which might change the world?

Photo by Markus Spiske on Unsplash

Photo by Daniel Schlodo on Unsplash

Our Dataset

- (interim) results and data from studies on Covid-19 vaccines from Pfizer/Biontech, Moderna, AstraZeneca and Janssen are publicly available
- We have selected the Gender and Age data as our star for the upcoming movie
- While the Gender information is readily available, age data requires some transformation for consistent data across all companies

Dive Into Details

Chat

What is your main insight?

Post into Chat

	Compound	Pfizer/Biontech	Moderna	AstraZeneca	Johnson & Johnson
Datasource					
Publication		New England Journal of Medicine	New England Journal of Medicine	The Lancet	FDA Briefing Document
Link		https://www.nejm.org/doi/full/10.1056/NEJMoa2034577	https://www.nejm.org/doi/full/10.1056/nejmoa2035389	https://www.sciencedirect.com/science/article/pii/S0140673620326611	https://www.fda.gov/media/146217/download
Vaccine		BNT162b2	mRBN-1273	AZD1222	Ad26.COV2.S
Pivotal Study on ClinicalTrials.gov		NCT04368728	NCT04470427	NCT04324606, NCT04400838, and NCT04444674	NCT04505722
Sex					
Male	Vaccine	9639	7923	2282	10924
Female	Vaccine	9221	7258	3525	8702
Undifferentiated	Vaccine				2
Unknown	Vaccine				2
Male	Placebo	9436	8062	2309	10910
Female	Placebo	9410	7108	3520	8777
Undifferentiated	Placebo				4
Unknown	Placebo				0
Age					
16 to 55 yr	Vaccine	10889		5089	
>55 yr	Vaccine	7971		494	
16 to 55 yr	Placebo	10896		5129	
>55 yr	Placebo	7950		480	
18 to < 65yr at risk	Vaccine		8888		
18 to < 65yr not at risk	Vaccine		2530		
>= 65 yr	Vaccine		3763		
18 to < 65yr at risk	Placebo		8886		
18 to < 65yr not at risk	Placebo		2535		
>= 65 yr	Placebo		3749		
18 to 59 yr	Vaccine				12830
>= 60 yr	Vaccine				6800
>= 65 yr	Vaccine				3984
18 to 59 yr	Placebo				12881
>= 60 yr	Placebo				6810
>= 65 yr	Placebo				4018

Spoiler: Key Idea of our Story

There is a **noticeable difference** for the demographic and baseline variables

Age and **Gender**

for the **AstraZeneca** vaccine

compared to the other studies

From Data into Insights

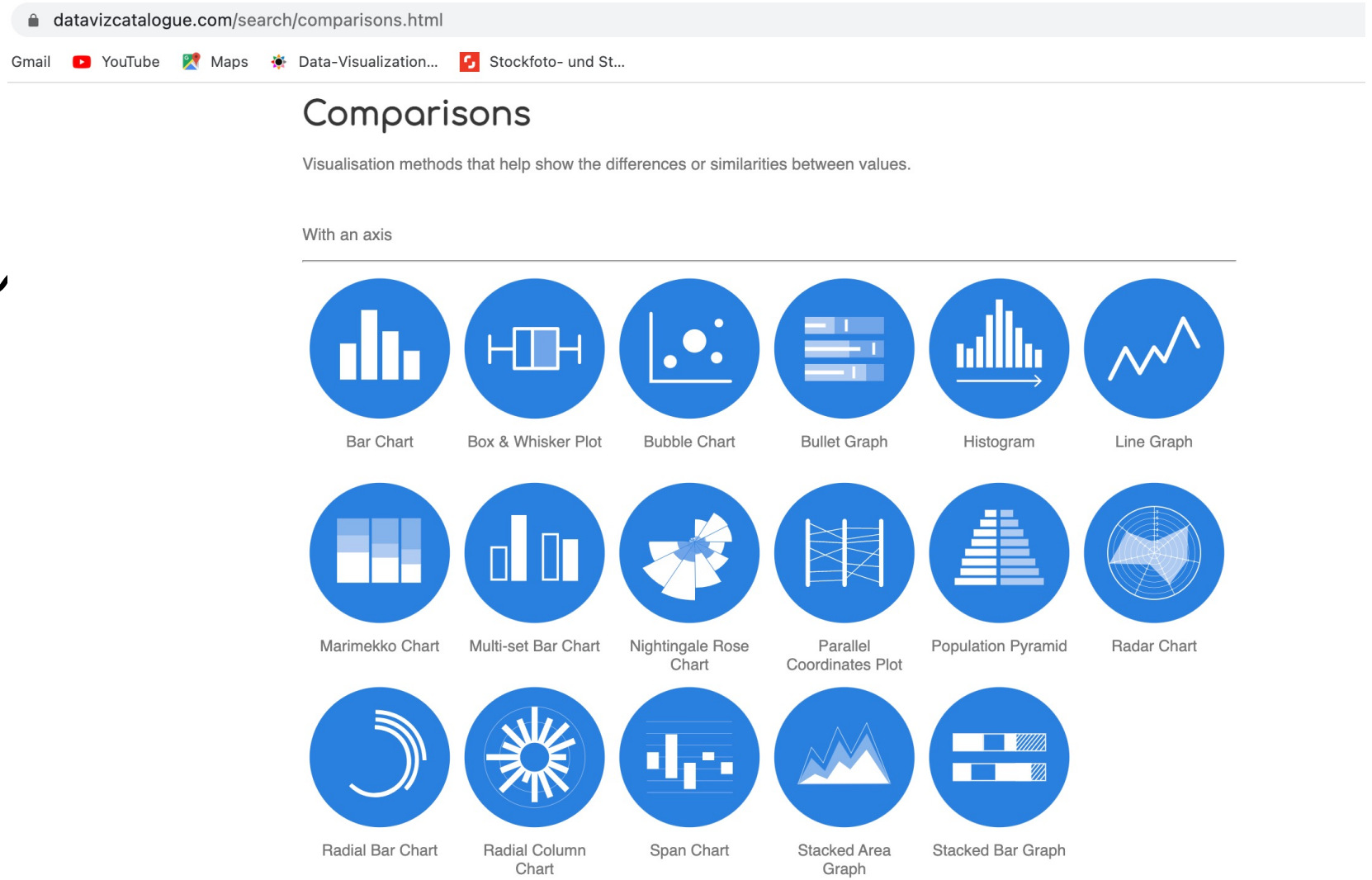
Following the exercise about the 3 minute story, we now try to create **powerful visuals** to tell our data story

Showing the Data

versus

Letting the Data tell a story

How to choose the best visual



www.datavizcatalogue.com

Choose the best visual for our data

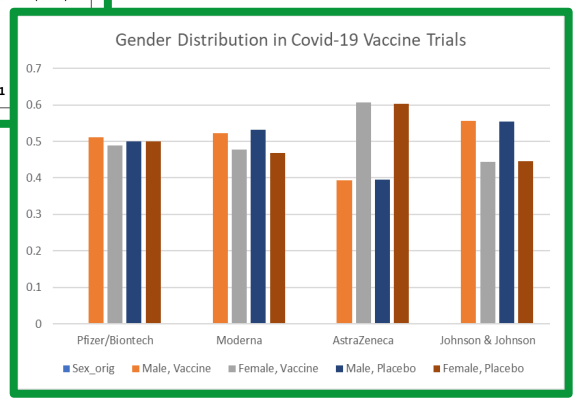
- President Macron called one vaccine in-effective **compared** to others
- What visual can we use to make the data tell us a story?

Table

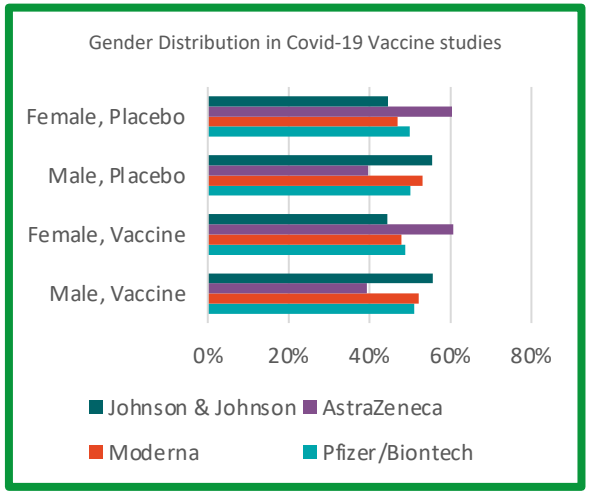
Table 1 - Gender distribution in Covid-19 vaccine studies

	Gender	Pfizer/Biontech		Moderna		AstraZeneca		Johnson & Johnson	
Covid19 Vaccine	Male	9.639	(51%)	7.923	(52%)	2.282	(39%)	10.924	(56%)
	Female	9.221	(49%)	7.258	(48%)	3.525	(61%)	8.702	(44%)
	Undifferentiated							2	(<1%)
	Unknown							2	(<1%)
	Total Sex	**18.860**	***(100%)***	**15.181**	***(100%)***	**5.807**	***(100%)***	**19.630**	***(100%)***
Placebo	Male	9.436	(50%)	8.062	(53%)	2.309	(40%)	10.910	(55%)
	Female	9.410	(50%)	7.108	(47%)	3.520	(60%)	8.777	
	Undifferentiated							4	
	Unknown								
	Total Sex	**18.846**	***(100%)***	**15.170**	***(100%)***	**5.829**	***(100%)***	**19.691**	

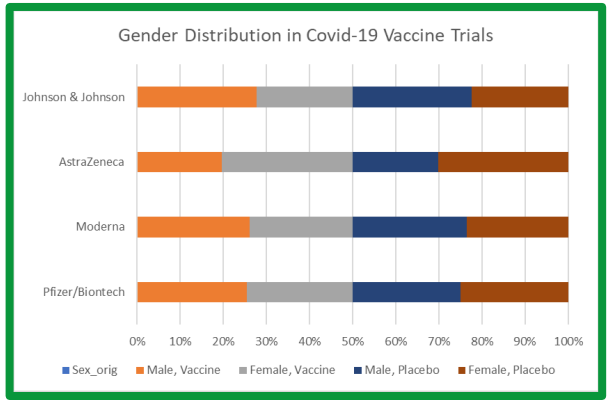
Vertical Bar Chart



Horizontal Bar Chart



Stacked Bar Chart



Table

What else could we improve?

Post into chat

Chat

bad

	Pfizer/ Biontech	Moderna	AstraZeneca	Johnson & Johnson
Male, Vaccine	9639	7923	2282	10924
Female, Vaccine	9221	7258	3525	8702
Male, Placebo	9436	8062	2309	10910
Female, Placebo	9410	7108	3520	8777

better

Drug	Gender	Pfizer/ Biontech		Moderna		AstraZeneca		Johnson & Johnson	
Vaccine	Male	9.639	51%	7.923	52%	2.282	39%	10.924	56%
	Female	9.221	49%	7.258	48%	3.525	61%	8.702	44%
Placebo	Male	9.436	50%	8.062	53%	2.309	40%	10.910	55%
	Female	9.410	50%	7.108	47%	3.520	60%	8.777	45%

Much better, but not best

Drug	Gender	Pfizer/Biontech	Moderna	AstraZeneca	Johnson & Johnson
Vaccine	Male	51%	52%	39%	56%
	Female	49%	48%	61%	44%
Placebo	Male	50%	53%	40%	55%
	Female	50%	47%	60%	45%

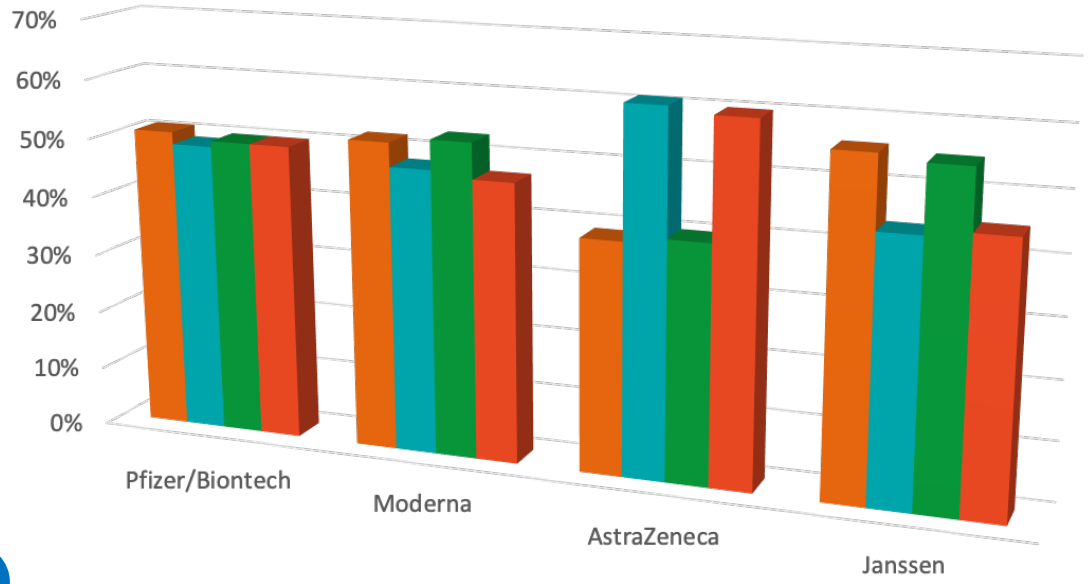
Bar Charts

What else could we improve?

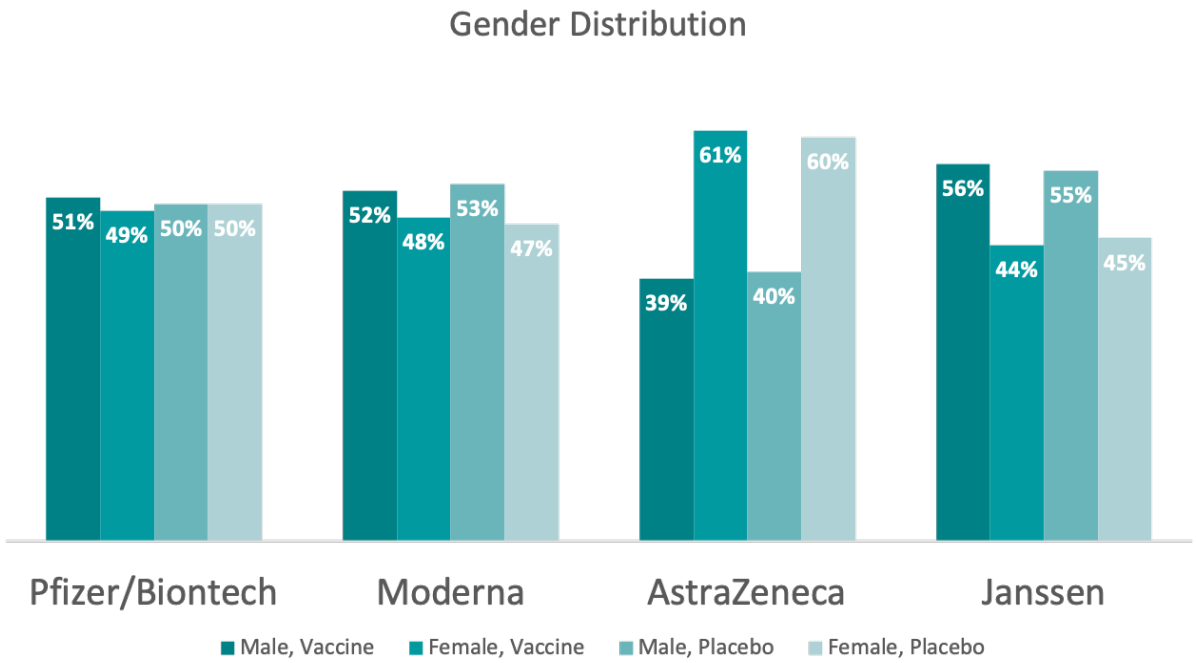
Post into chat

Chat

Really bad



Better



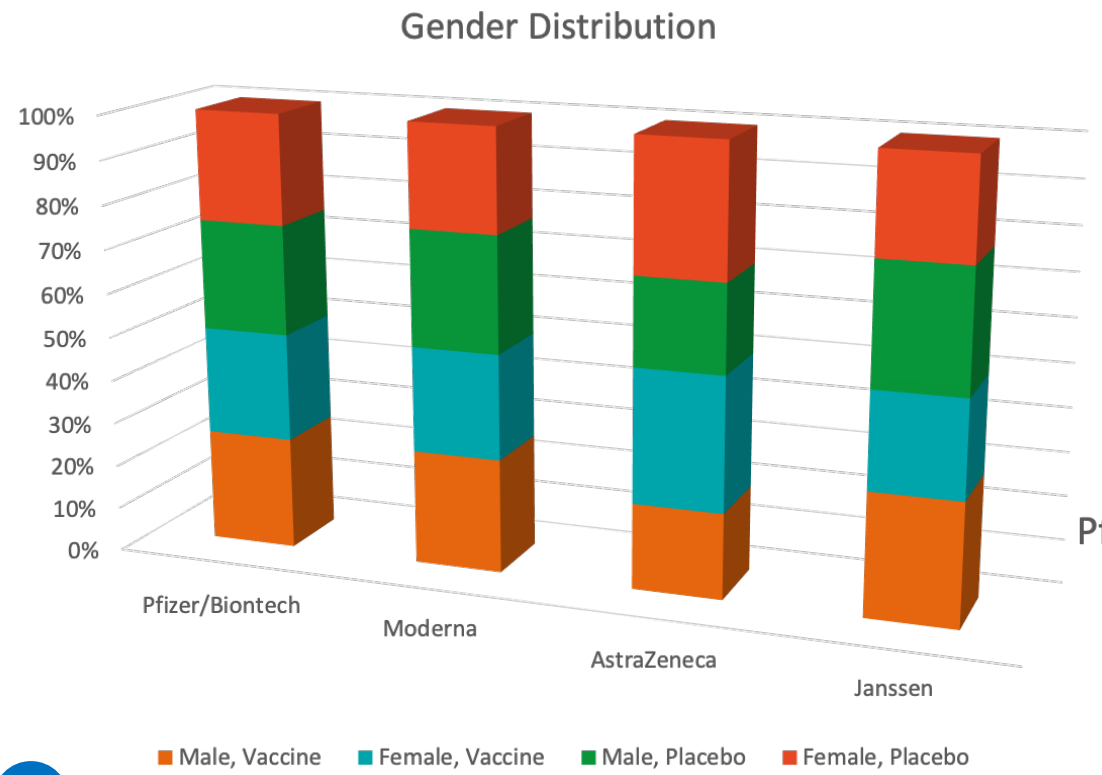
Stacked Bar Chart

What else could we improve?

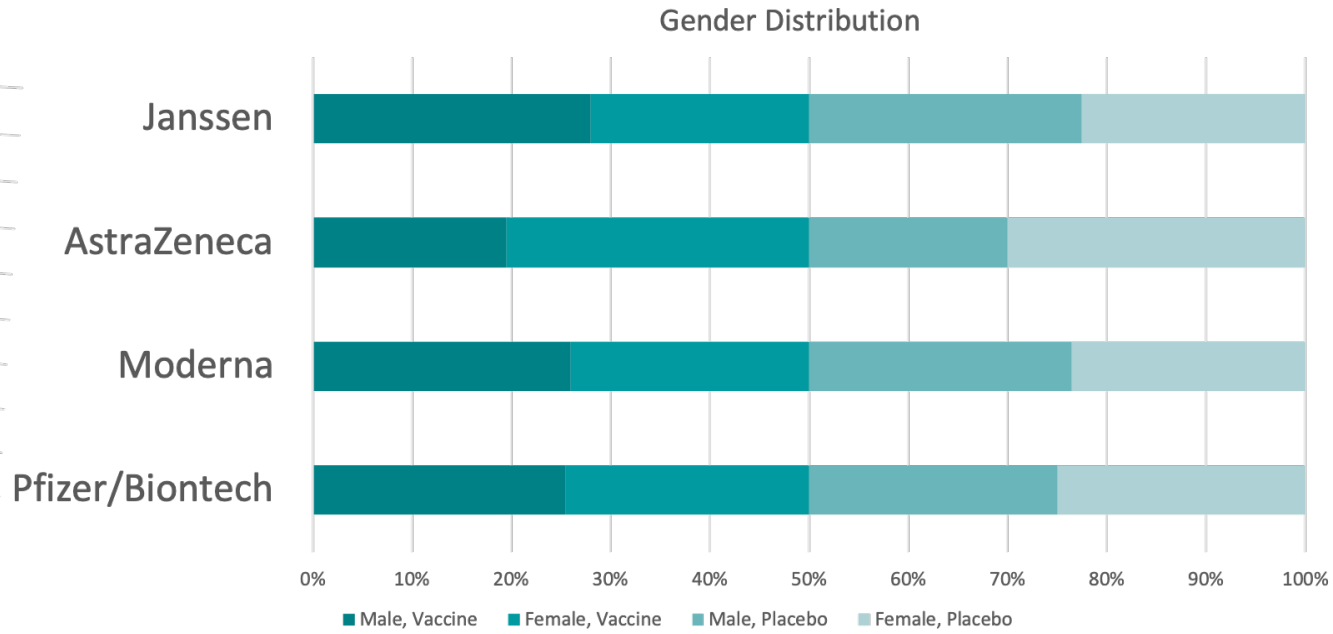
Chat

Post into chat

Really bad



Better



Chat

Interactive

Which of these examples do you think are best to use now and why?

Post it in the chat

M

Our Choice: Stacked Bar Chart

We want to compare the **relative size** of different ratios among the four companies. Stacked bar chart appears to offer these options

M

Side Story: Data Engineering

- Studies use different age ranges, which makes it difficult to compare
- Age has to be normalized to allow a comparison across companies
- www.ourworldindata.org uses age ranges 0-14 (children), 15 to 64 (working population), 65 and older (elderly)
- People older than 55 consists of about 45% between 55 and 65 and 55% older than 65.
- **Mapping** “15 to 55 years” + 0.45*”>55 years” to build “15 to 64 years”, “>55 years”*0.55 to build “65 years and older”

	Compound	Pfizer/Biontech	Moderna	AstraZeneca	Johnson & Johnson
Age					
16 to 55 yr	Vaccine	10889		5089	
>55 yr	Vaccine	7971		494	
16 to 55 yr	Placebo	10896		5129	
>55 yr	Placebo	7950		480	
18 to < 65yr at risk	Vaccine		8888		
18 to < 65yr not at risk	Vaccine		2530		
>= 65 yr	Vaccine		3763		
18 to < 65yr at risk	Placebo		8886		
18 to < 65yr not at risk	Placebo		2535		
>= 65 yr	Placebo		3749		
18 to 59 yr	Vaccine				12830
>= 60 yr	Vaccine				6800
>= 65 yr	Vaccine				3984
18 to 59 yr	Placebo				12881
>= 60 yr	Placebo				6810
>= 65 yr	Placebo				4018

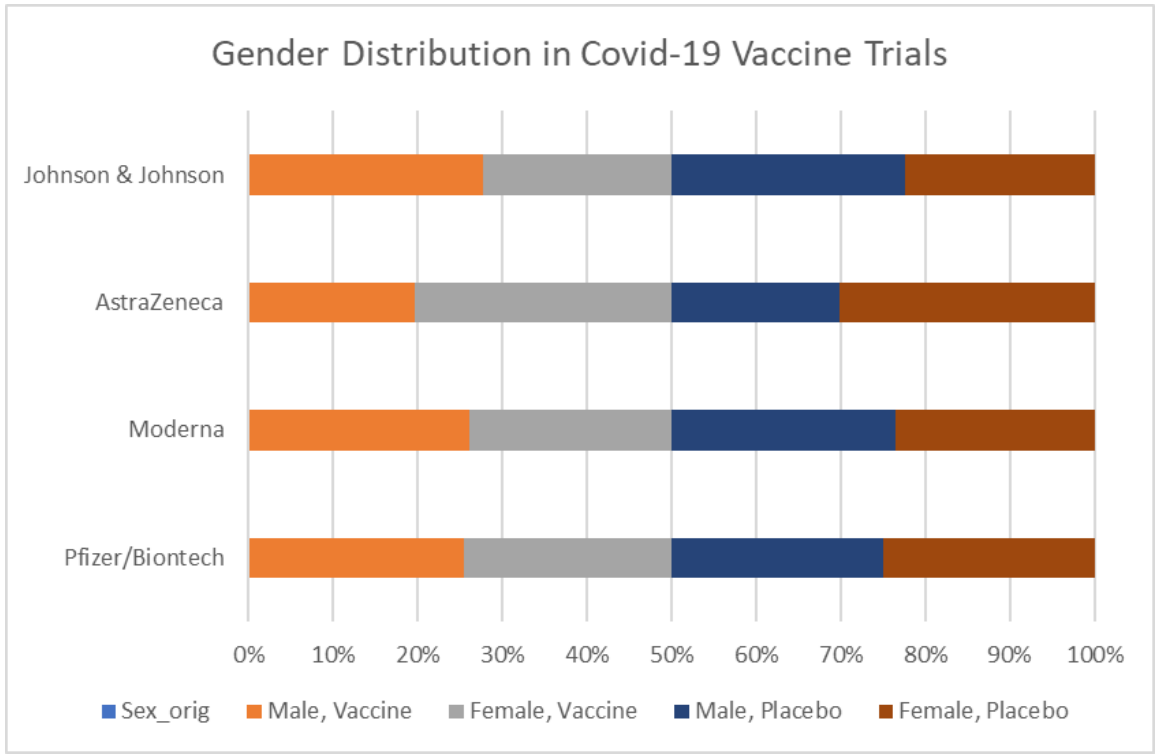
Mapping

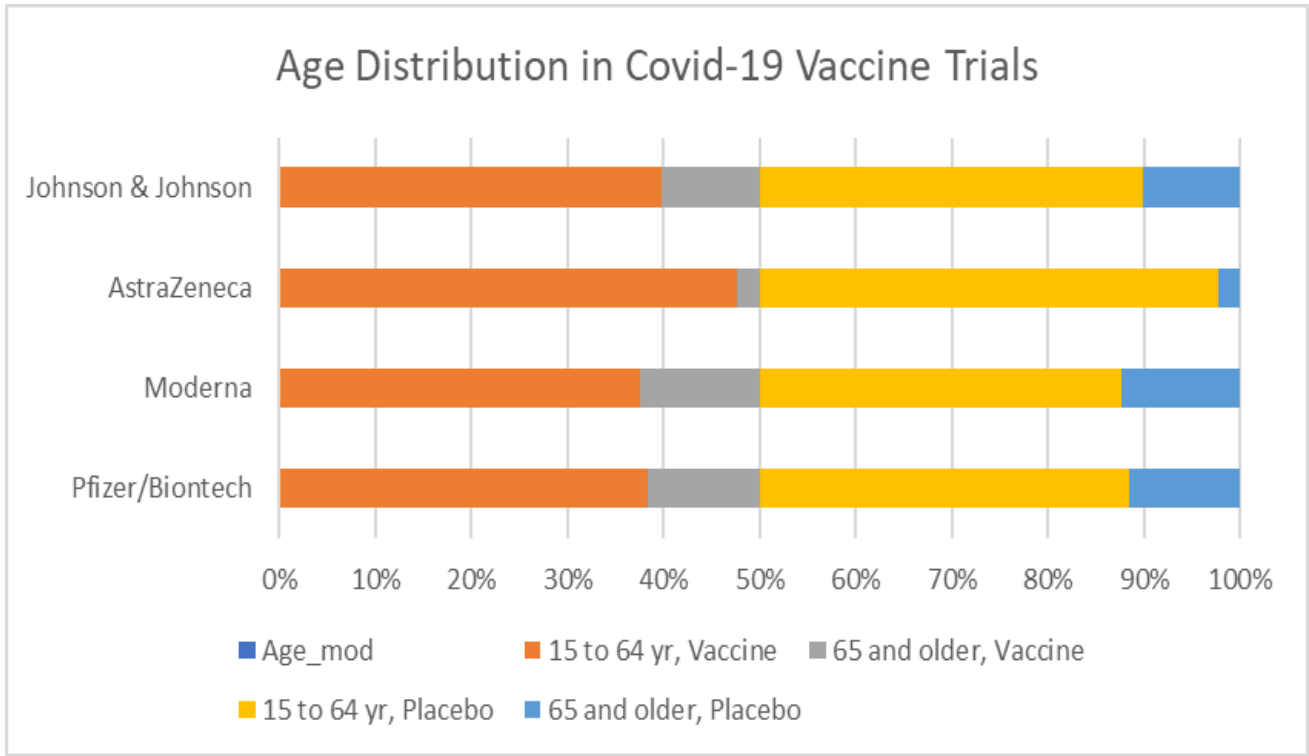
	Compound	Pfizer/Biontech	Moderna	AstraZeneca	Johnson & Johnson
Age					
15 to 64 years	Vaccine	14476	11418	5311	15646
>65 years	Vaccine	4384	3763	272	3984
15 to 64 years	Placebo	14474	11421	5345	15673
>65 years	Placebo	4373	3749	264	4018

Effective Visual

Showing the Data
versus
Letting the Data tell a story

How can we modify the charts to support our story most effectively?





Special Effects

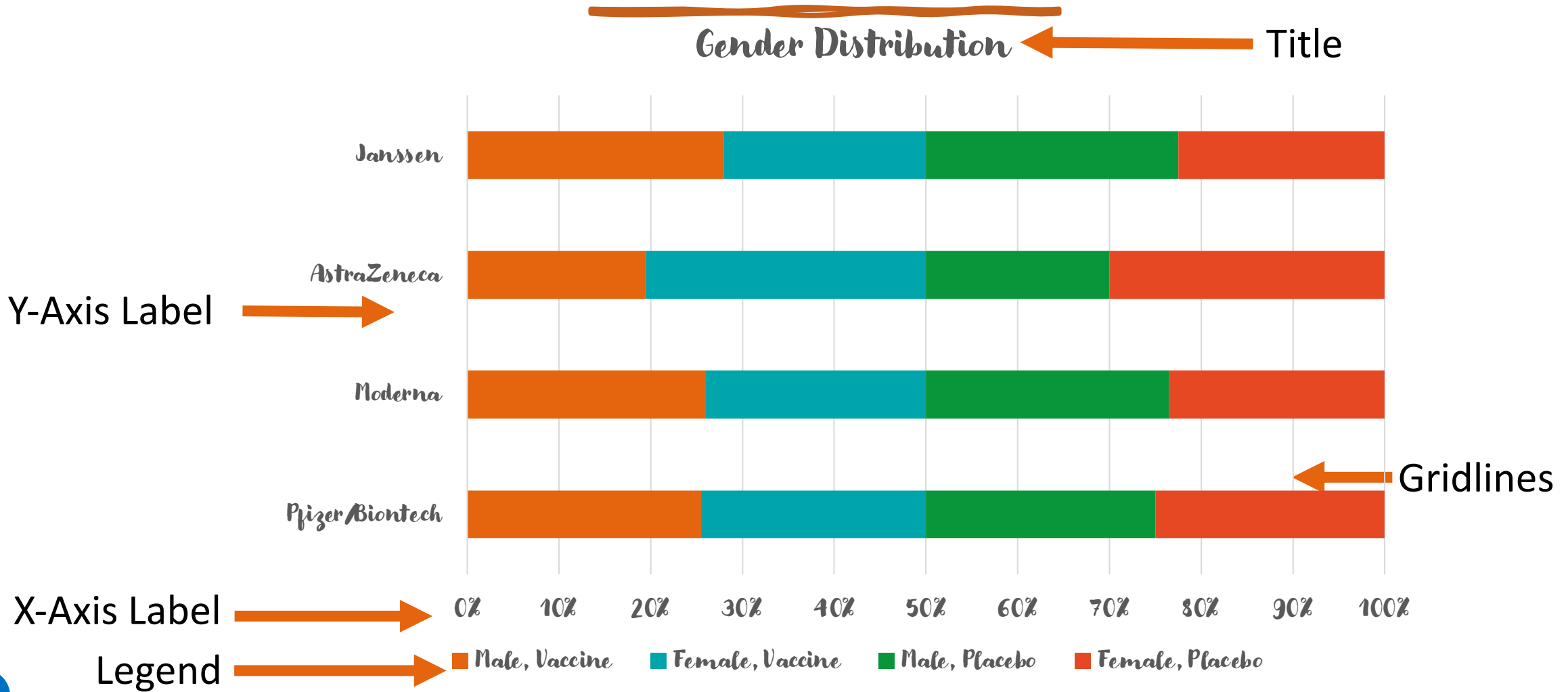
Reduce the clutter

- Figure contains a lot of elements
- Do we need them all? Or could we modify these?

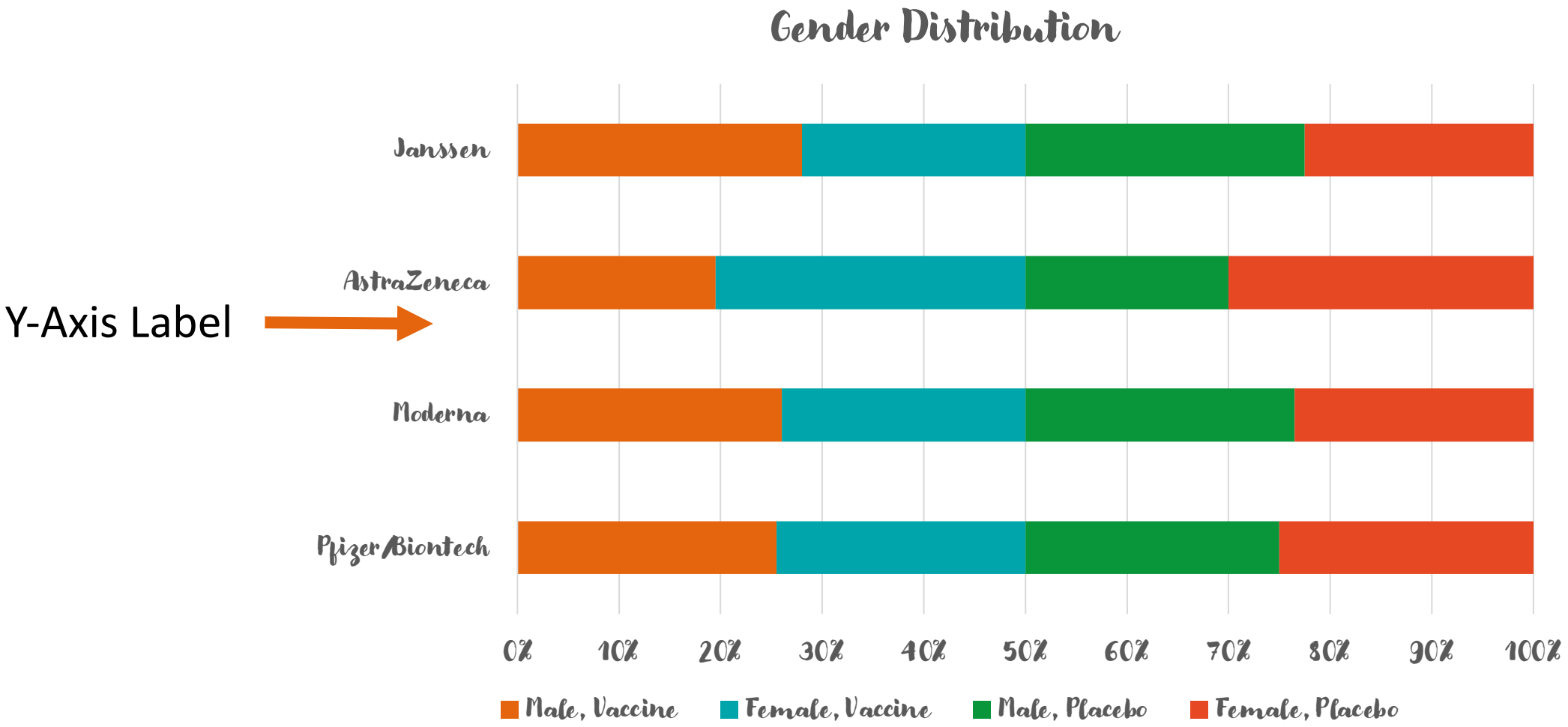
Focus attention of the audience

- We want to control what the audience focusses on
- How can we achieve this?

SFX: Reduce the Clutter

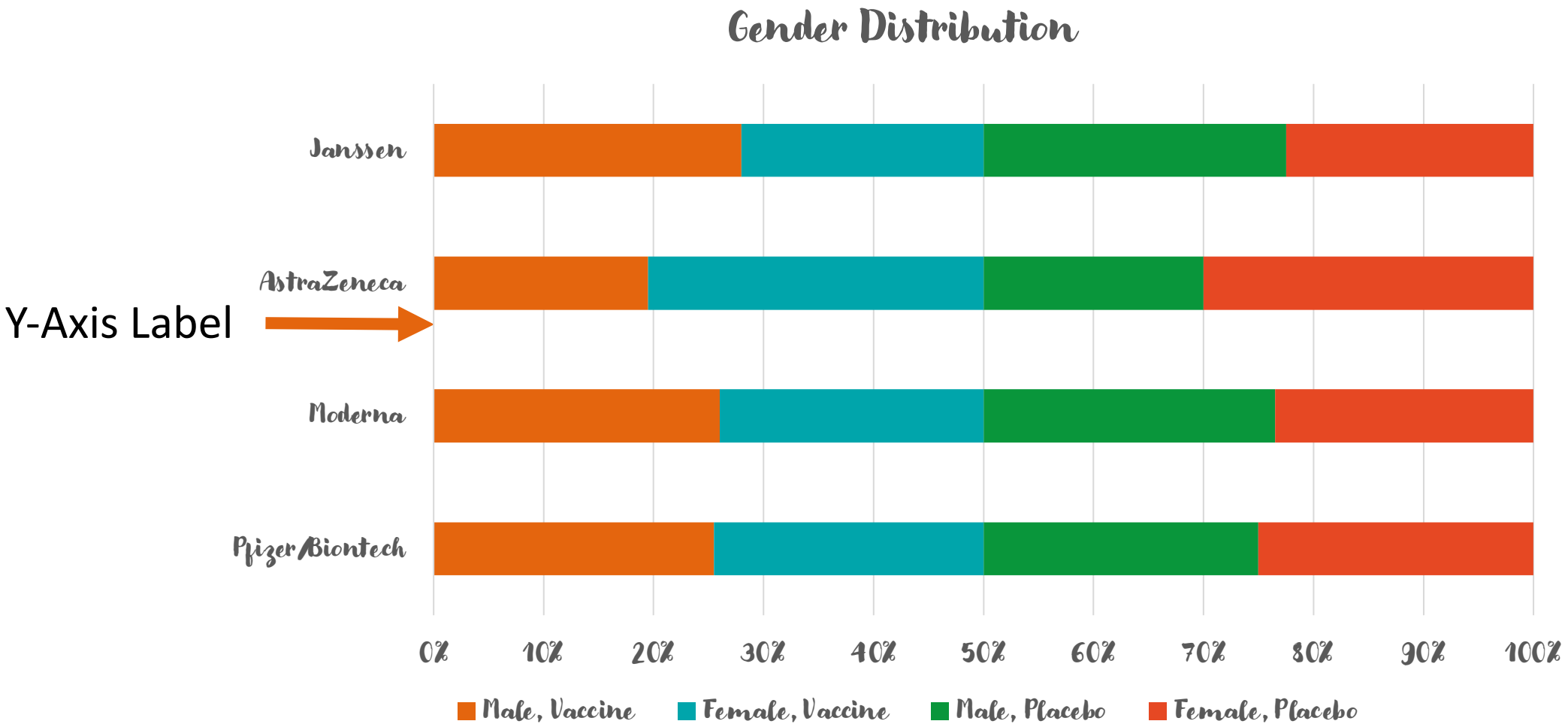


SFX: Reduce the Clutter

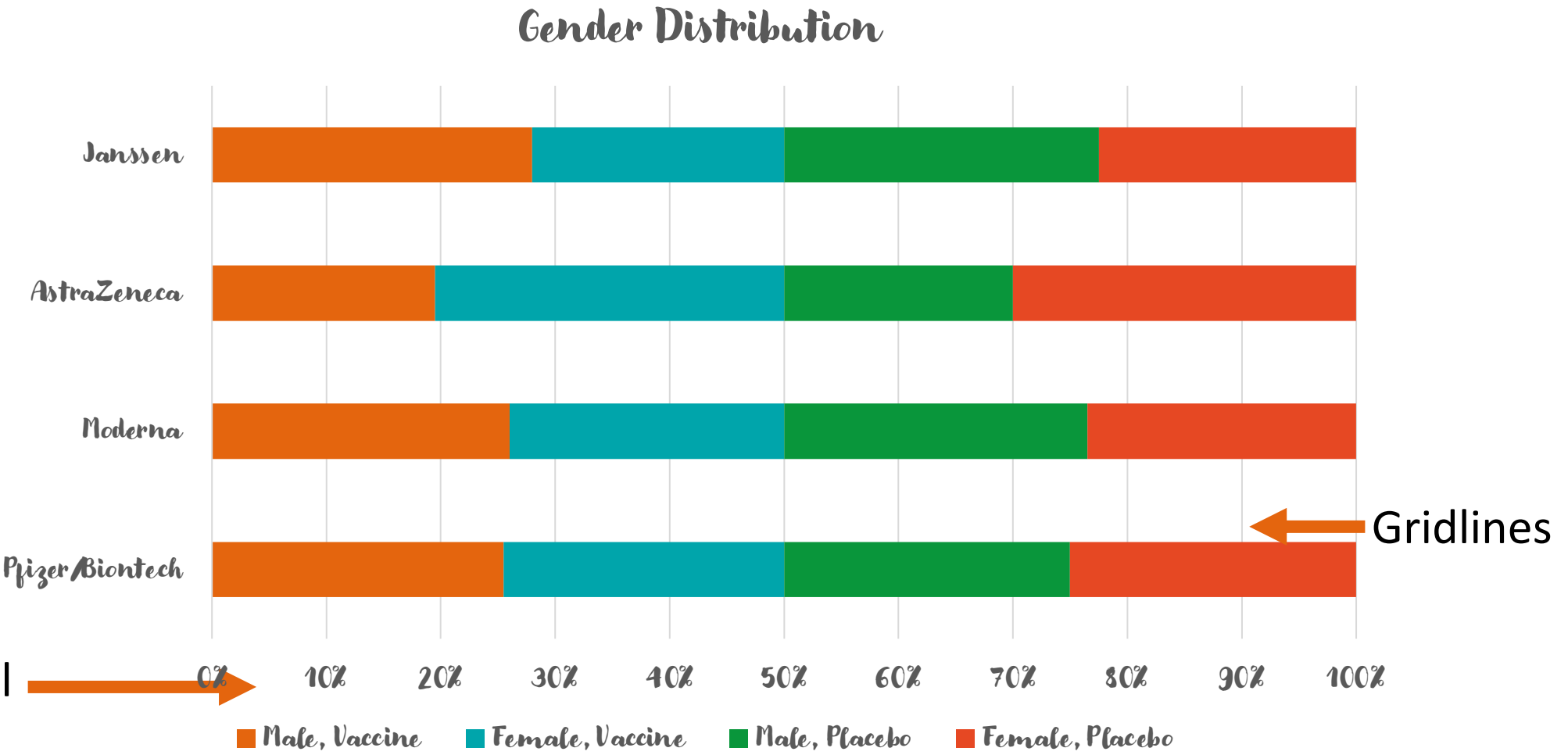


Y-Axis Label

SFX: Reduce the Clutter



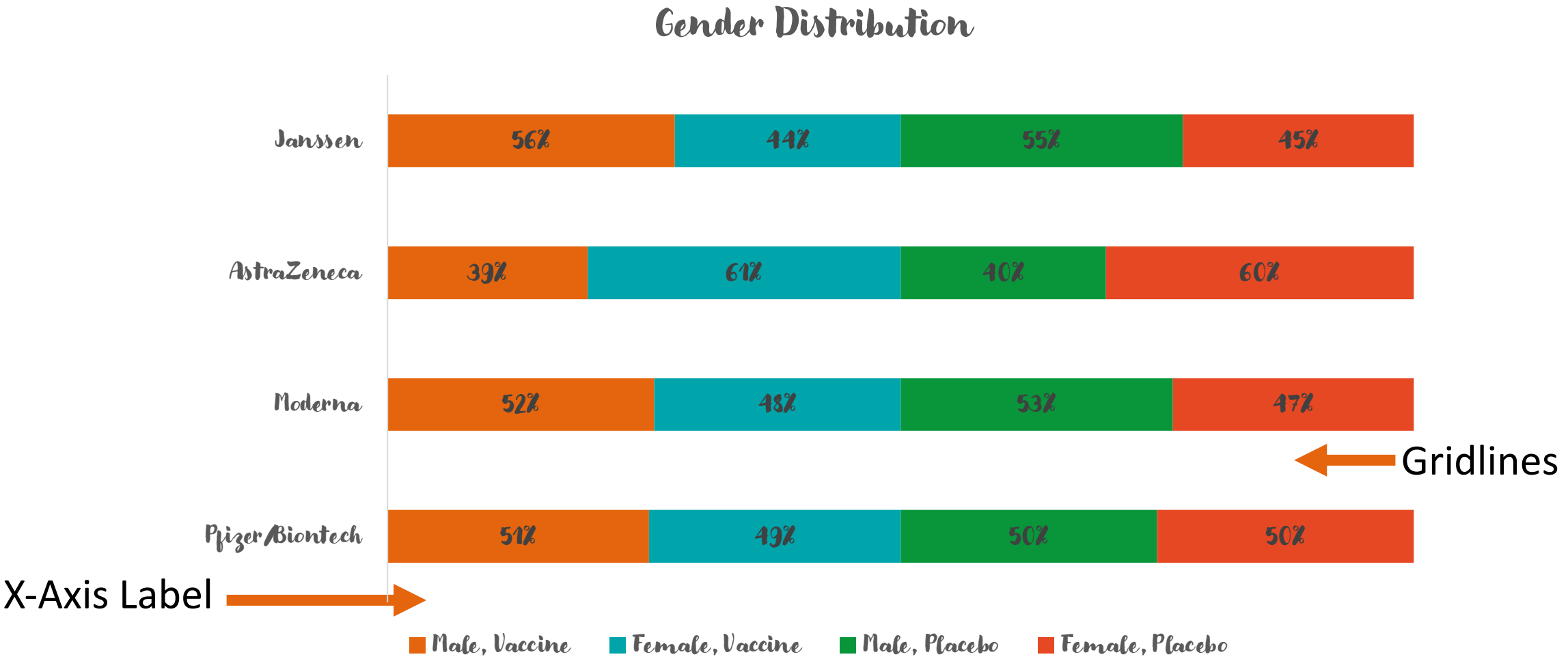
SFX: Reduce the Clutter



Gridlines

X-Axis Label

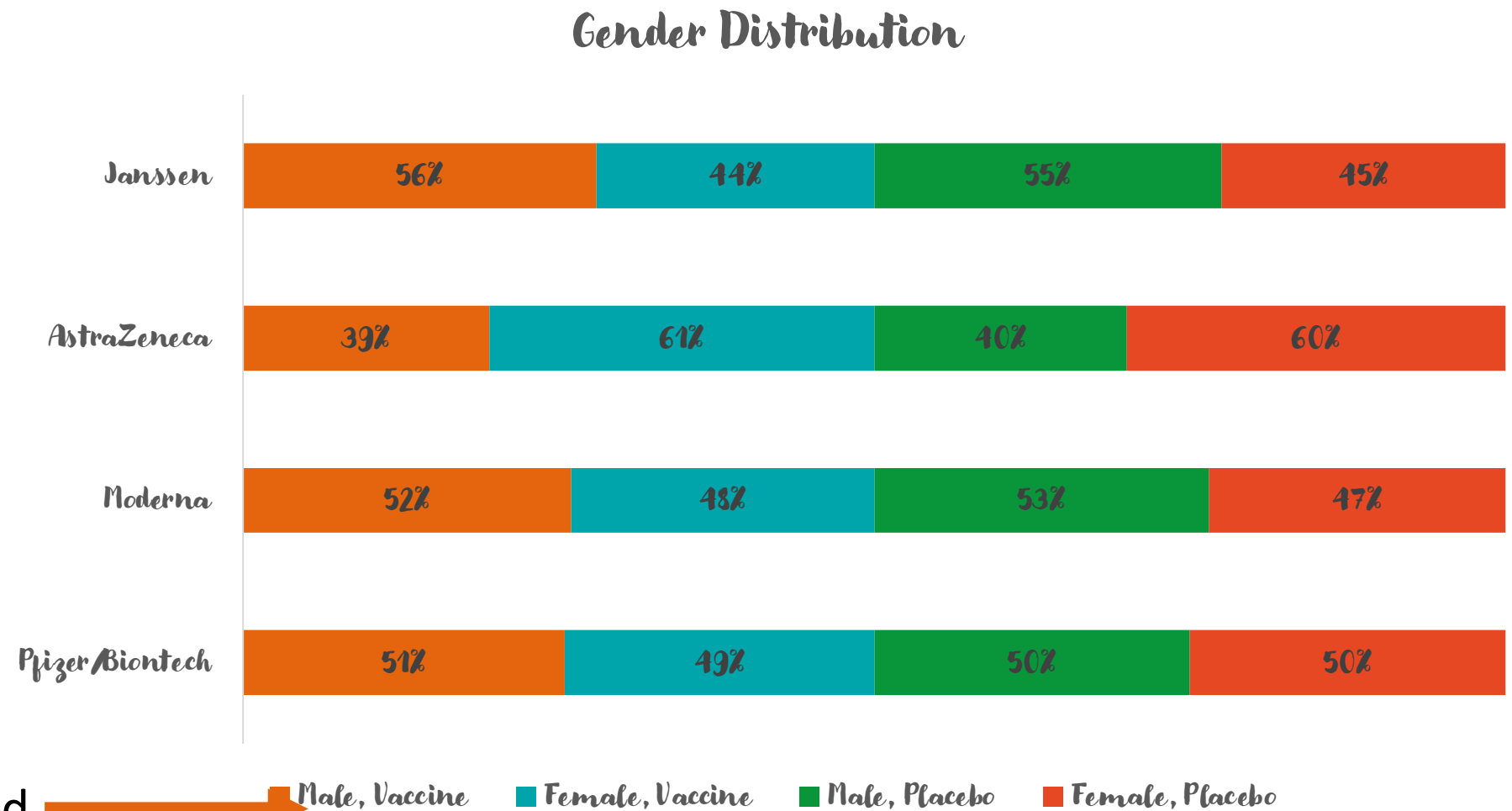
SFX: Reduce the Clutter



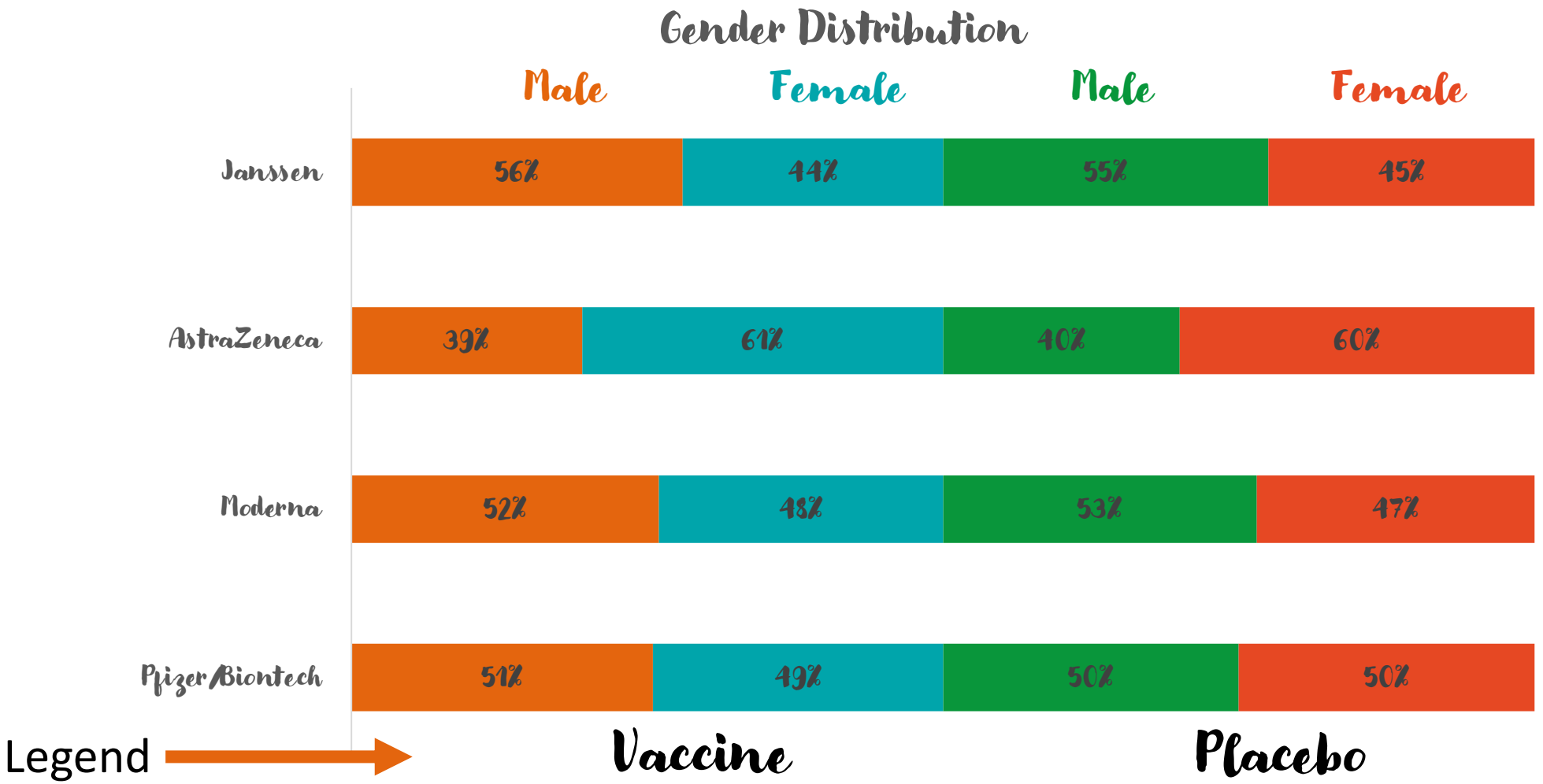
Gridlines

X-Axis Label

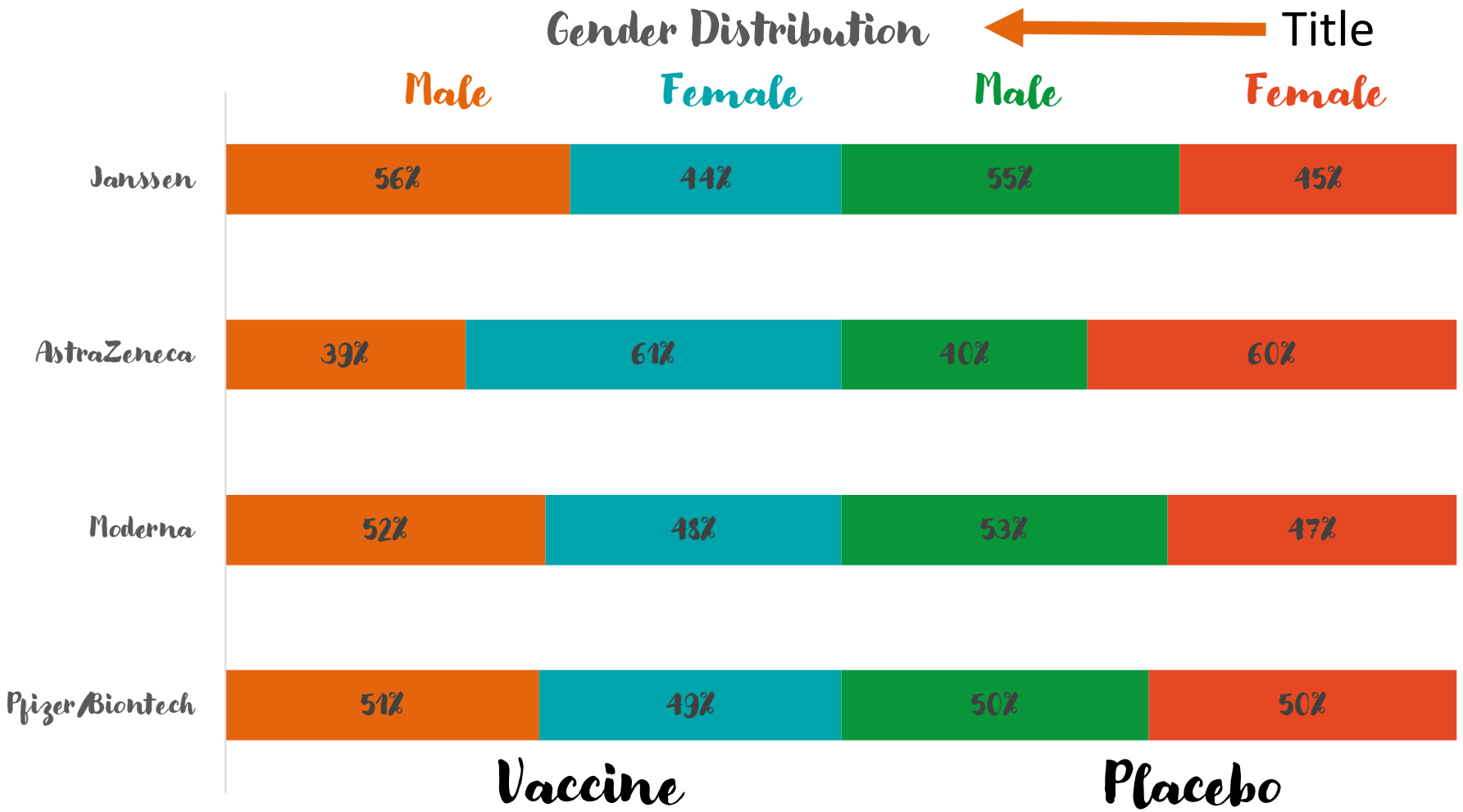
SFX: Reduce the Clutter



SFX: Reduce the Clutter



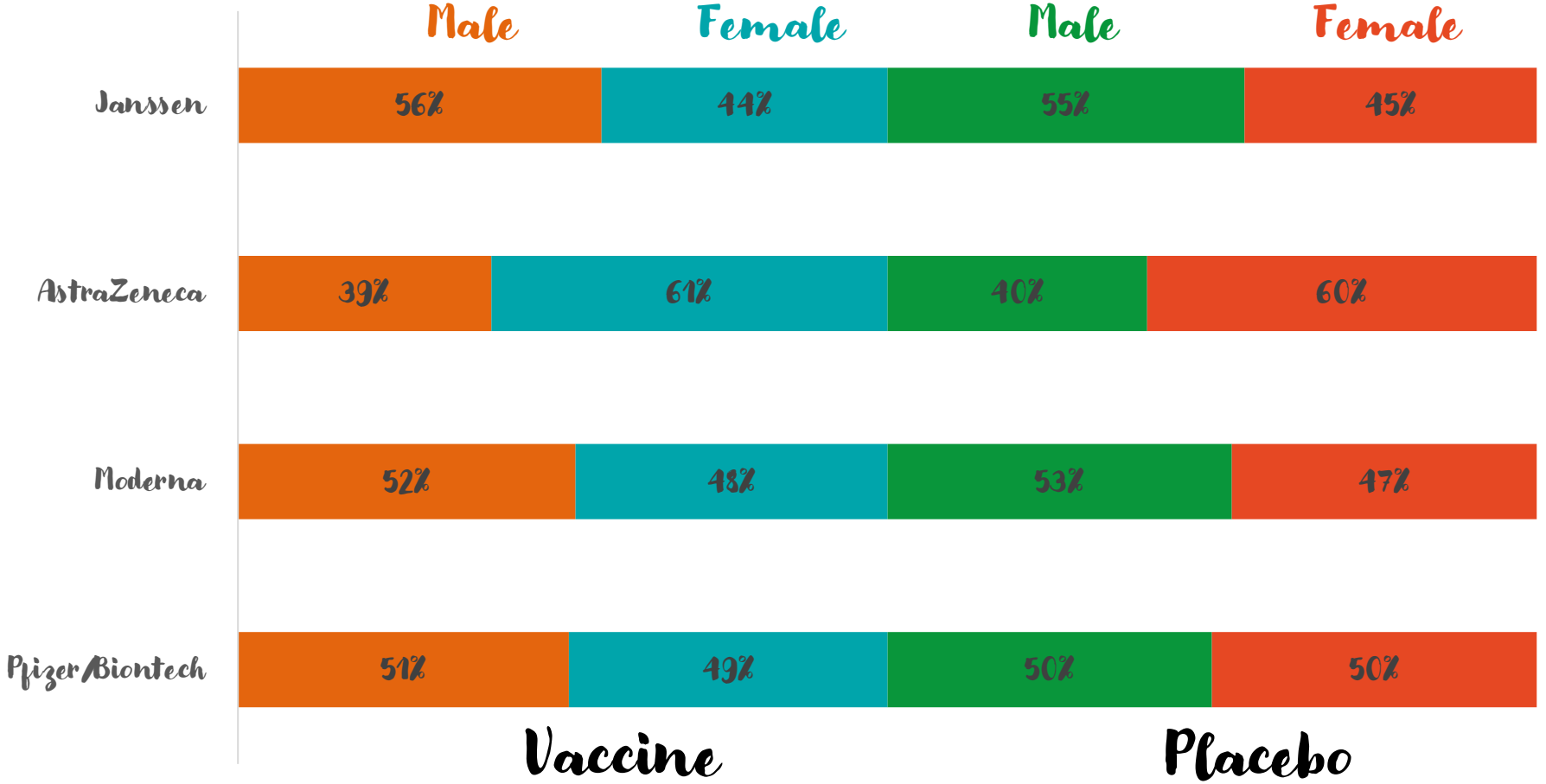
SFX: Reduce the Clutter



SFX: Reduce the Clutter

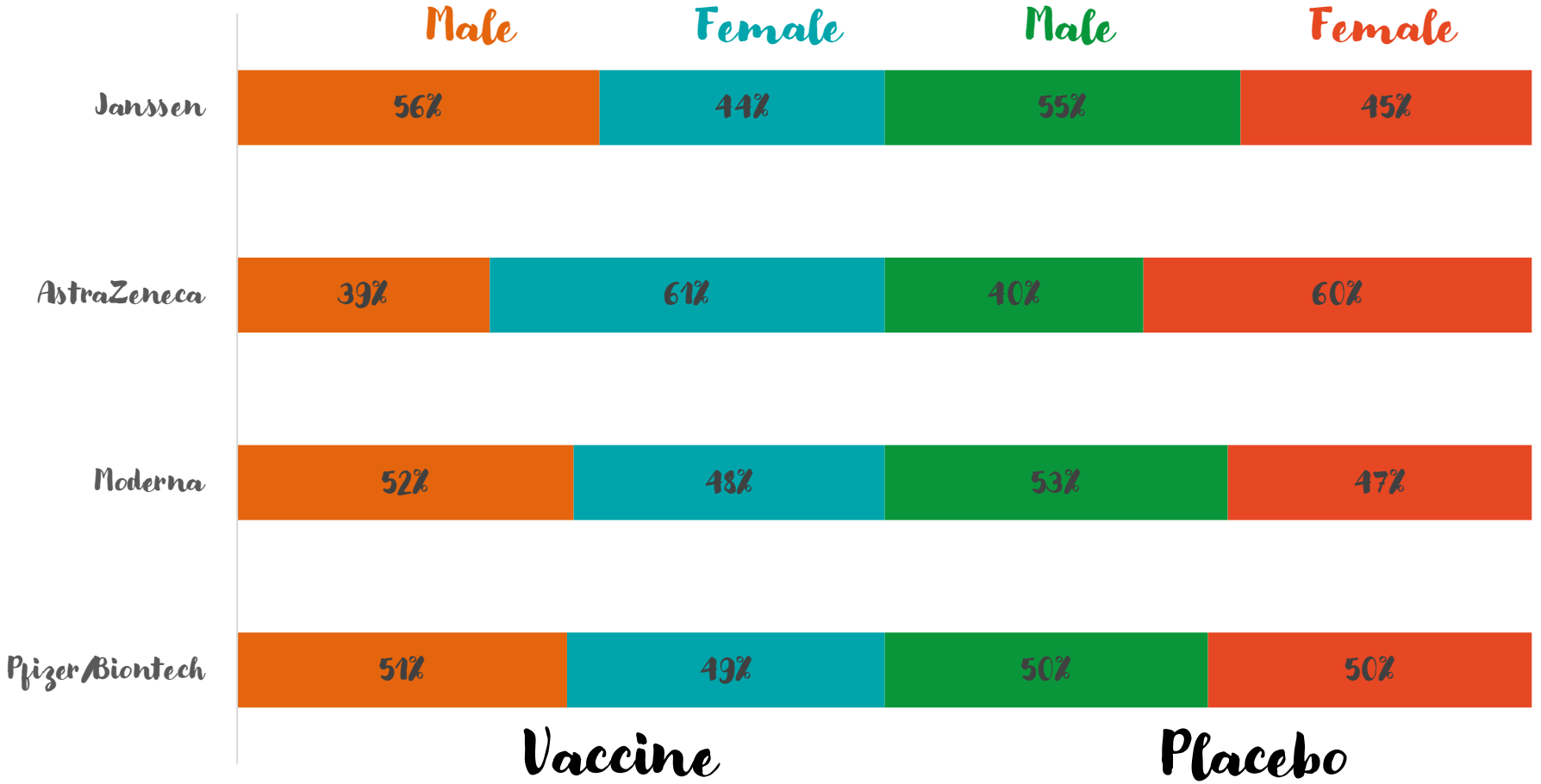
Gender Distribution in COVID-19 Vaccine studies

Title



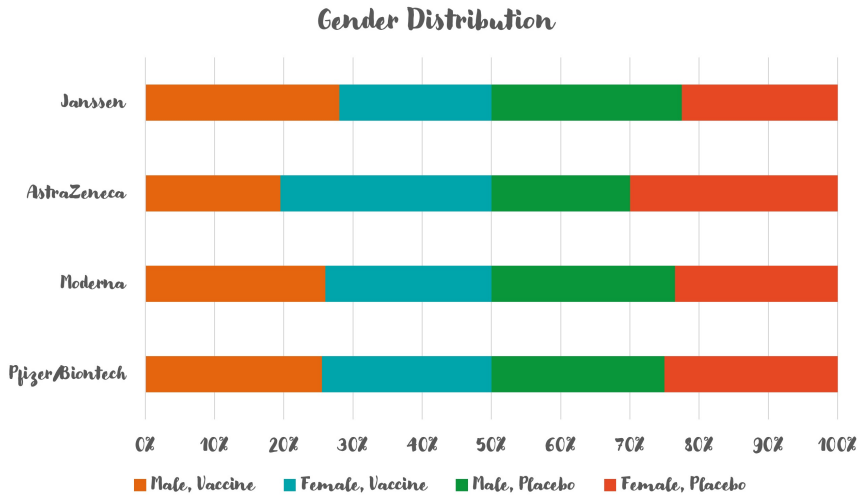
SFX: Reduce the Clutter

Gender Distribution in COVID-19 Vaccine studies

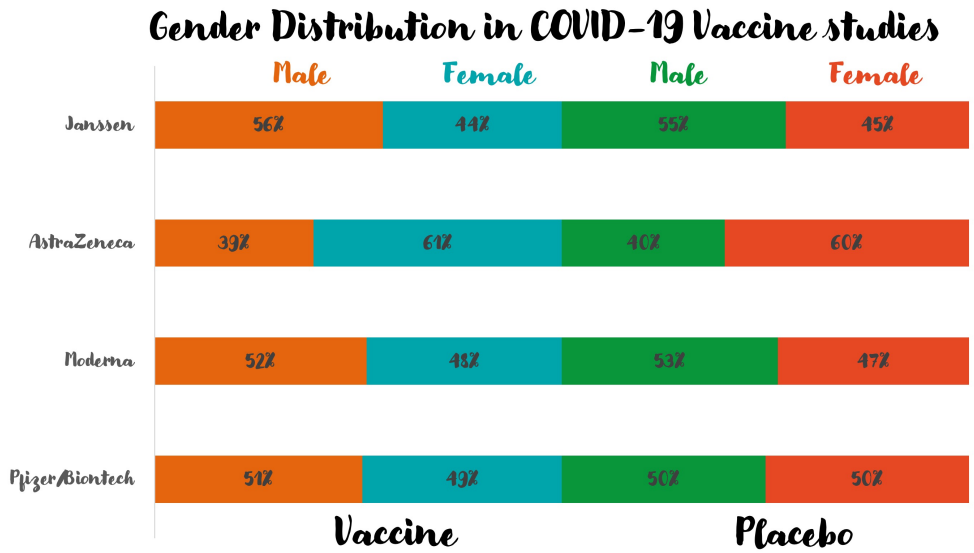


SFX: Reduce the Clutter

before

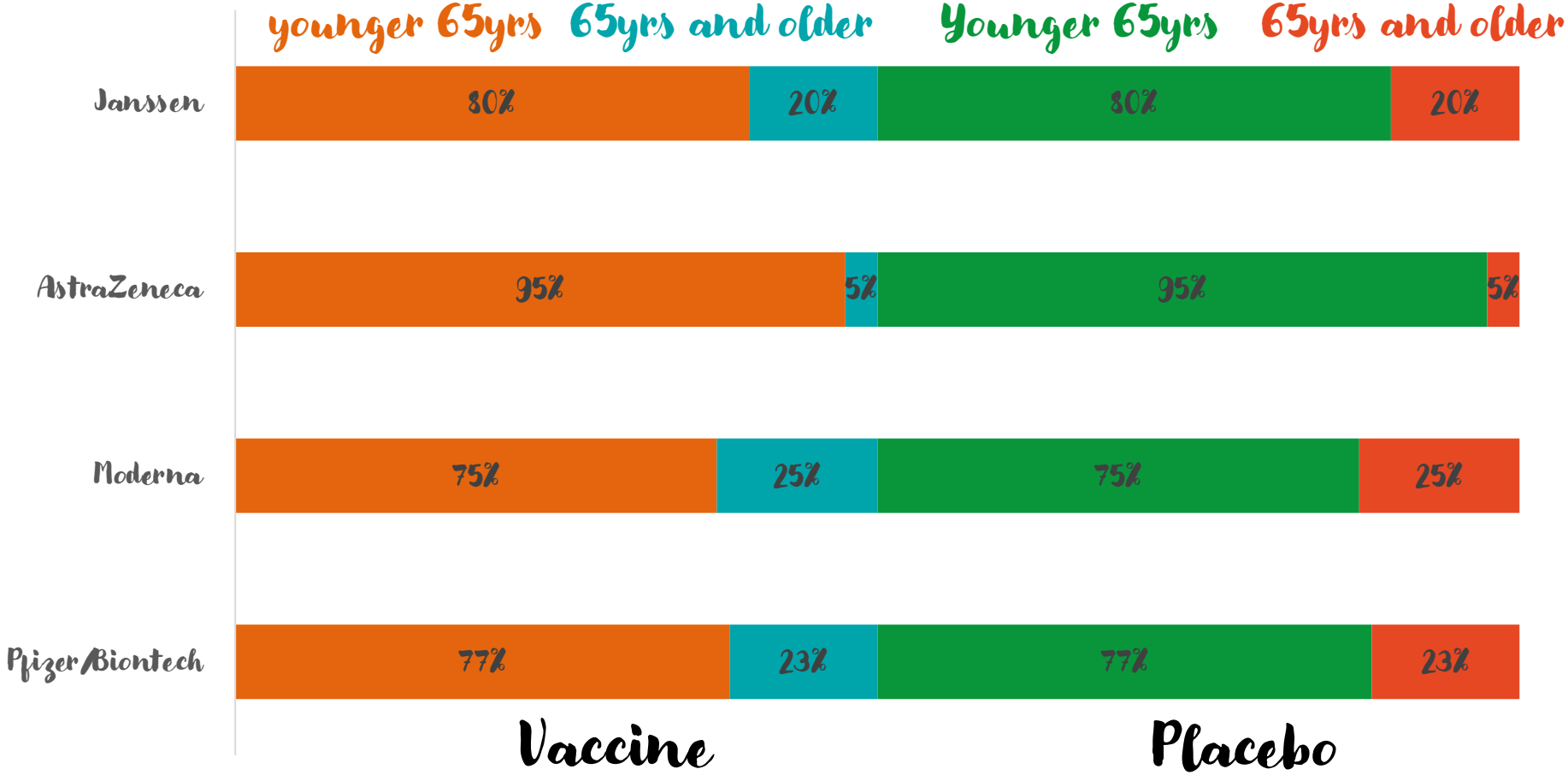


after



SFX: Reduce the Clutter

Age Distribution in COVID-19 Vaccine studies



SFX: Focus Attention

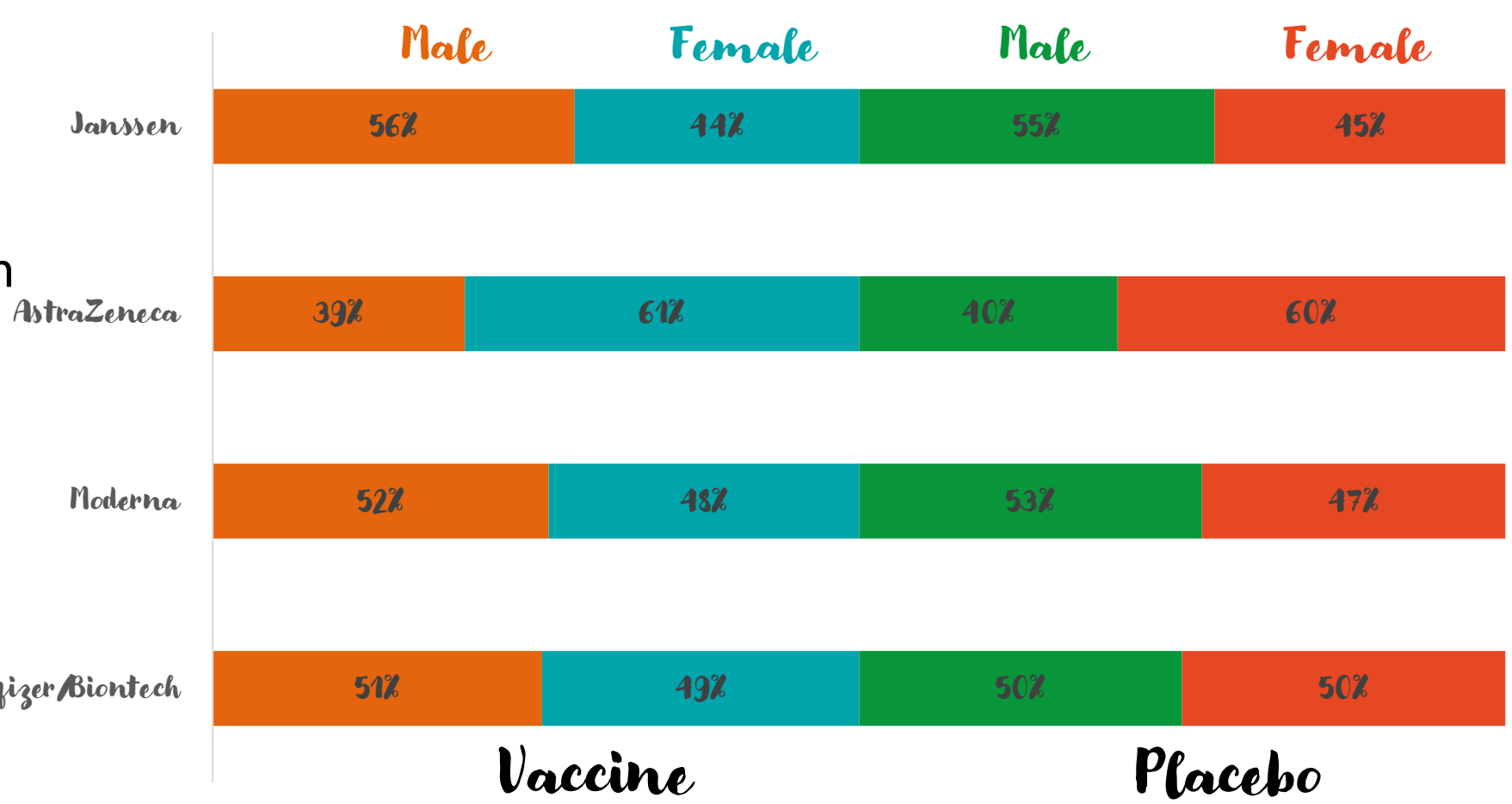
Exercise: How Many 6s do you see?

9638476018622784653578

9**6**3847**6**018**6**22784**6**53578

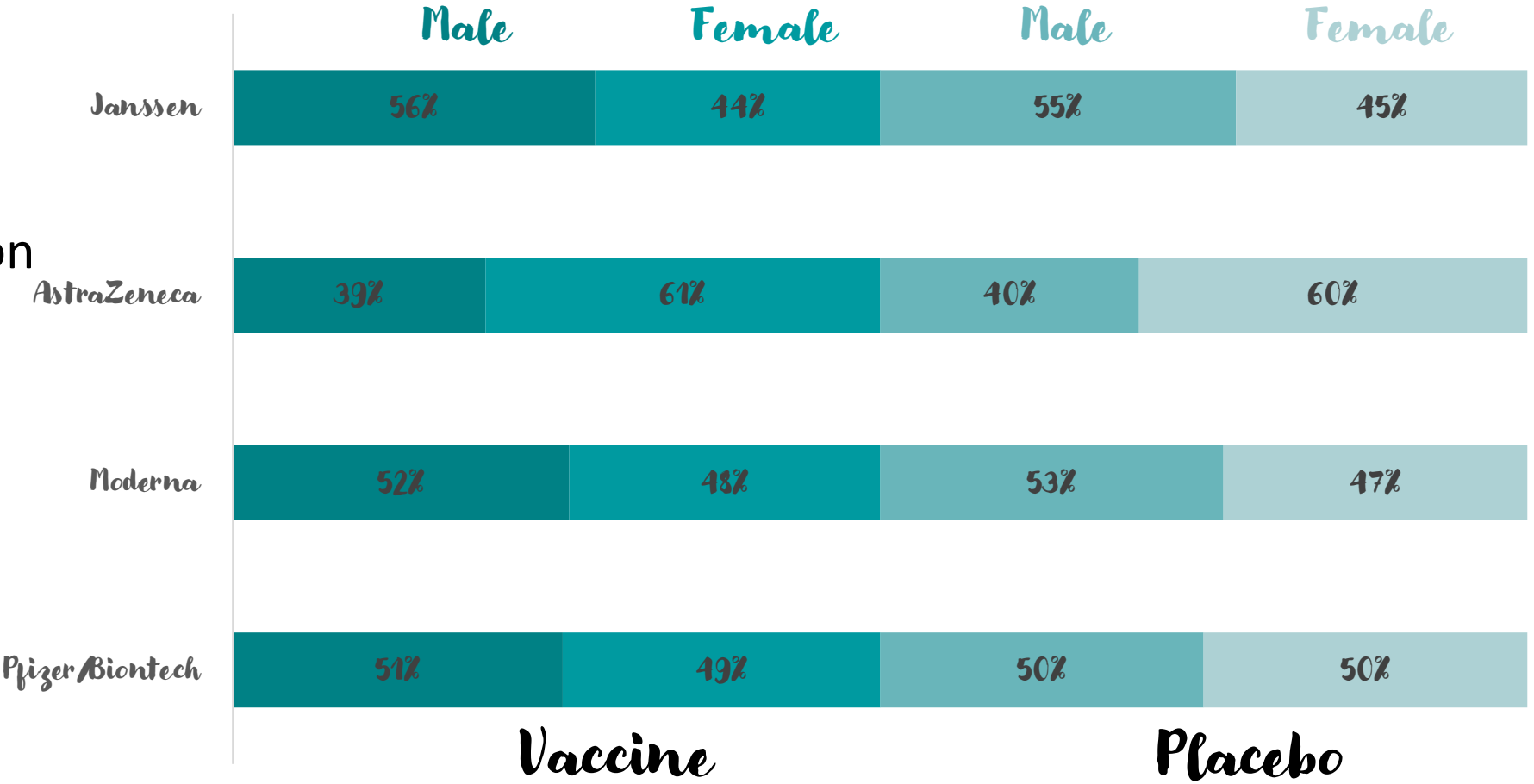
SFX: Focus Attention

- Color
- Fonts
- Style
- All Data?
- Comparison
- Annotate



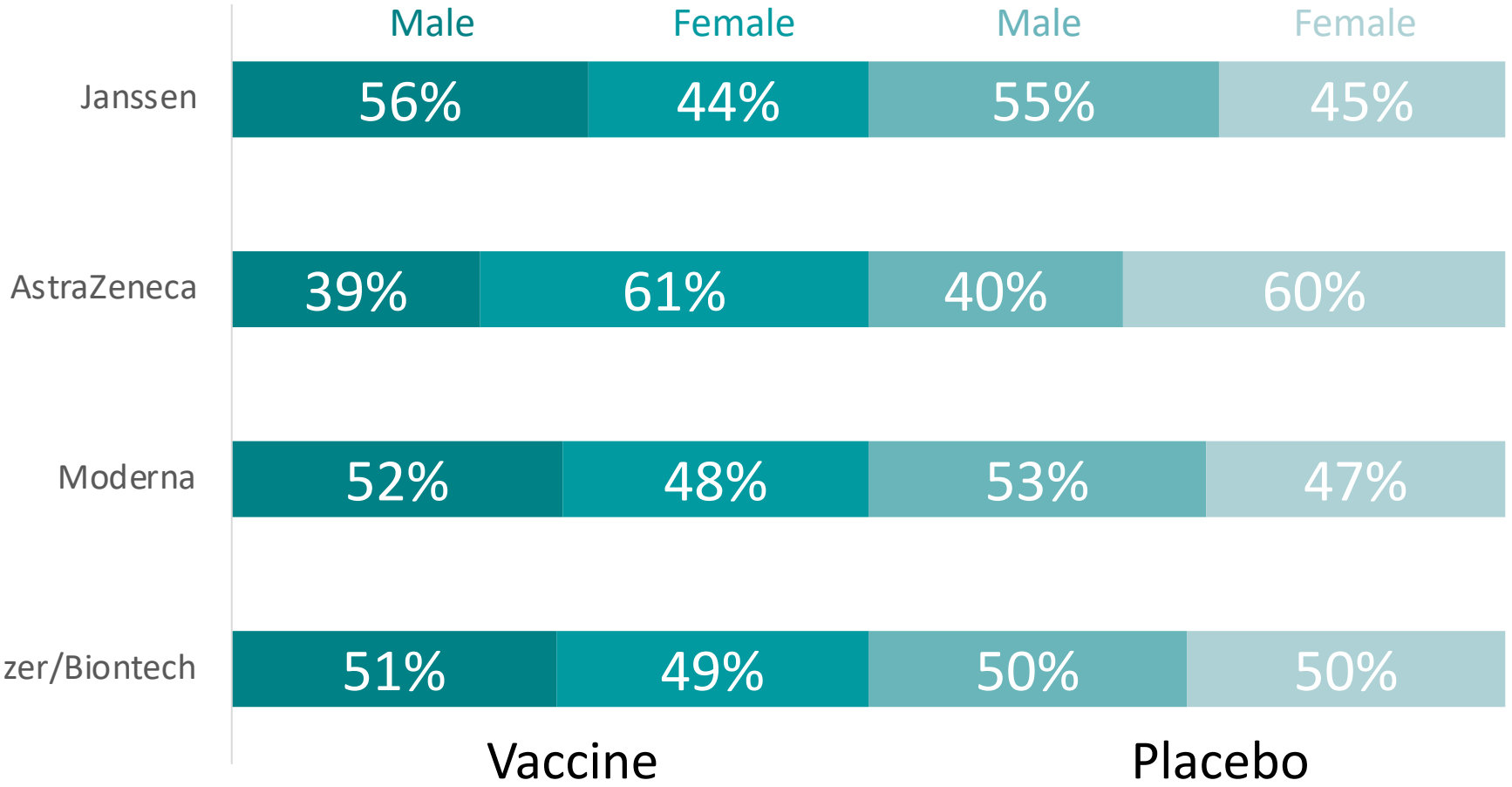
SFX: Focus Attention

- ✓ Color
- ○ Fonts
- ○ Style
- ○ All Data?
- ○ Comparison
- ○ Annotate



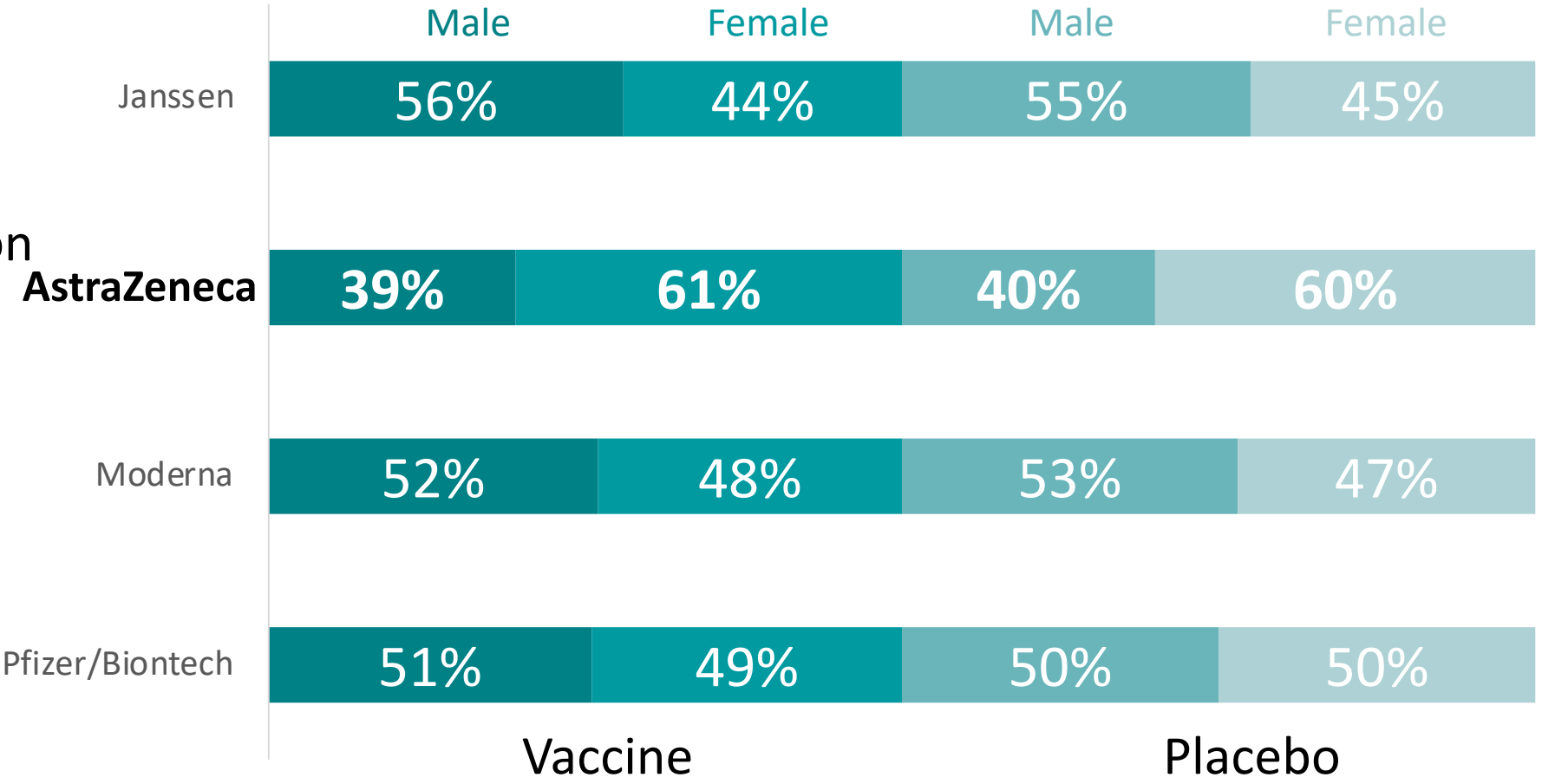
SFX: Focus Attention

- ✓ Color
- ✓ Fonts
- ○ Style
- ○ All Data?
- ○ Comparison
- ○ Annotate



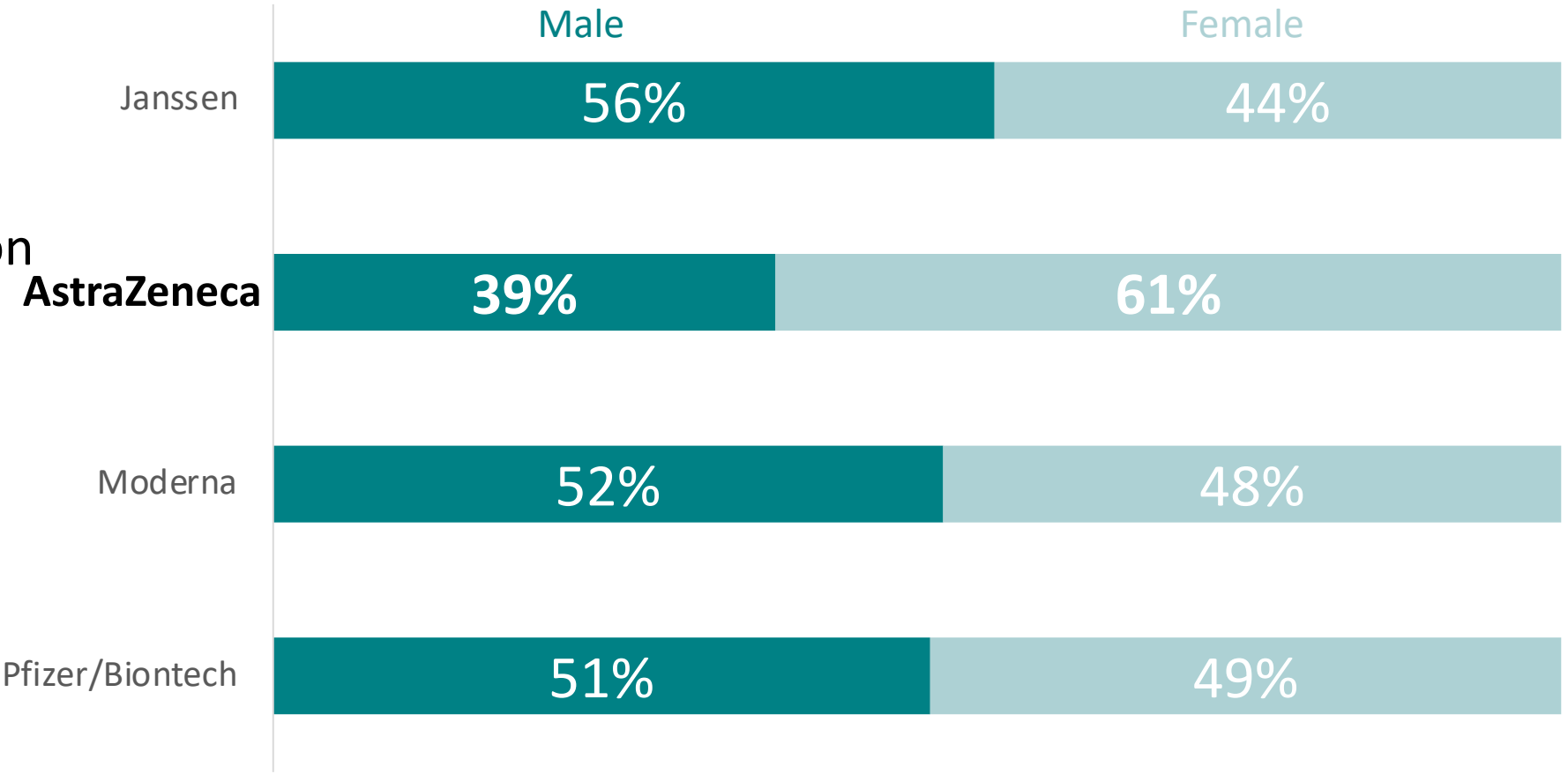
SFX: Focus Attention

- ✓ Color
- ✓ Fonts
- ✓ Style
- ○ All Data?
- ○ Comparison
- ○ Annotate



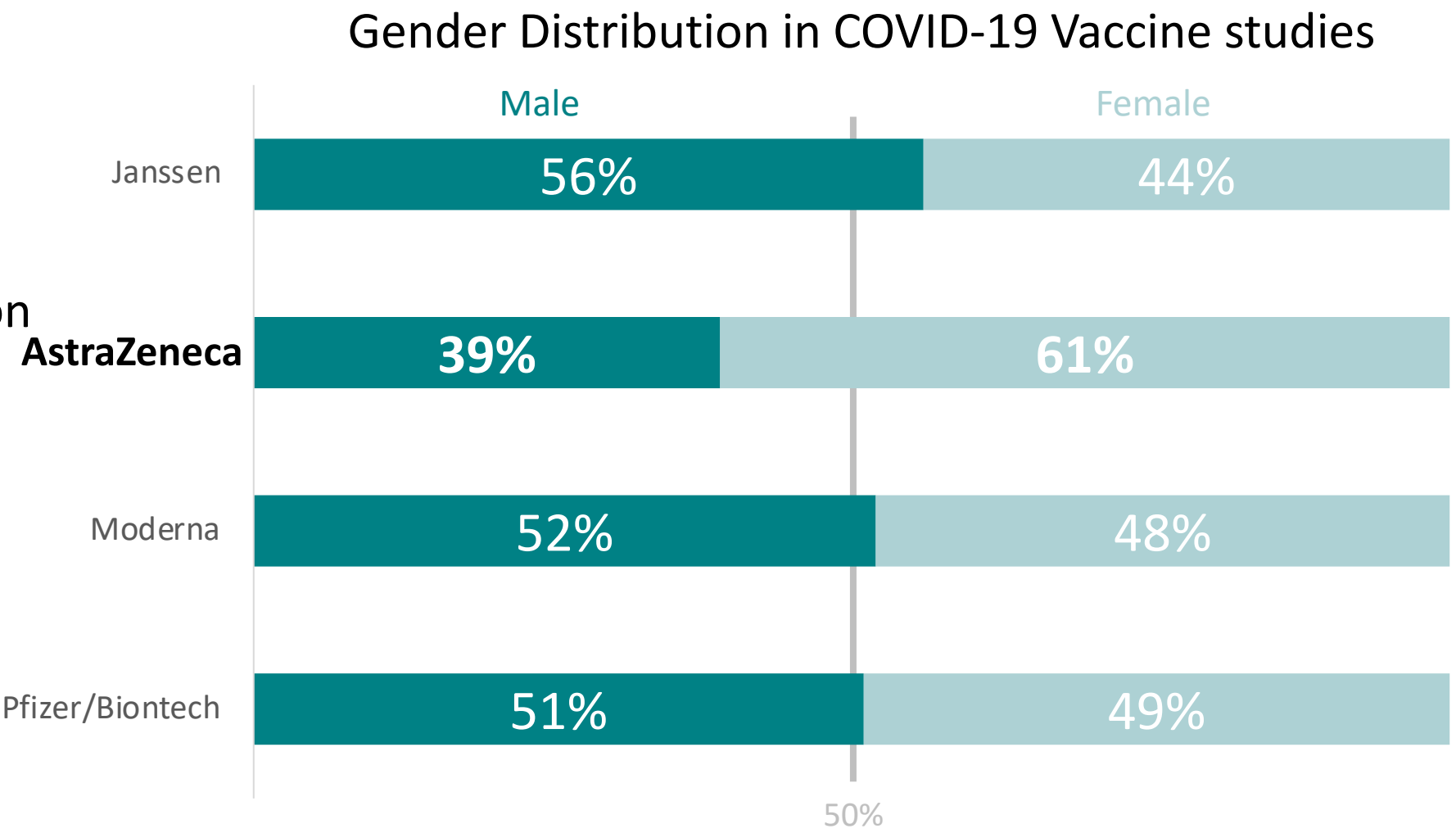
SFX: Focus Attention

- ✓ Color
- ✓ Fonts
- ✓ Style
- ✓ All Data?
- ○ Comparison
- ○ Annotate



SFX: Focus Attention

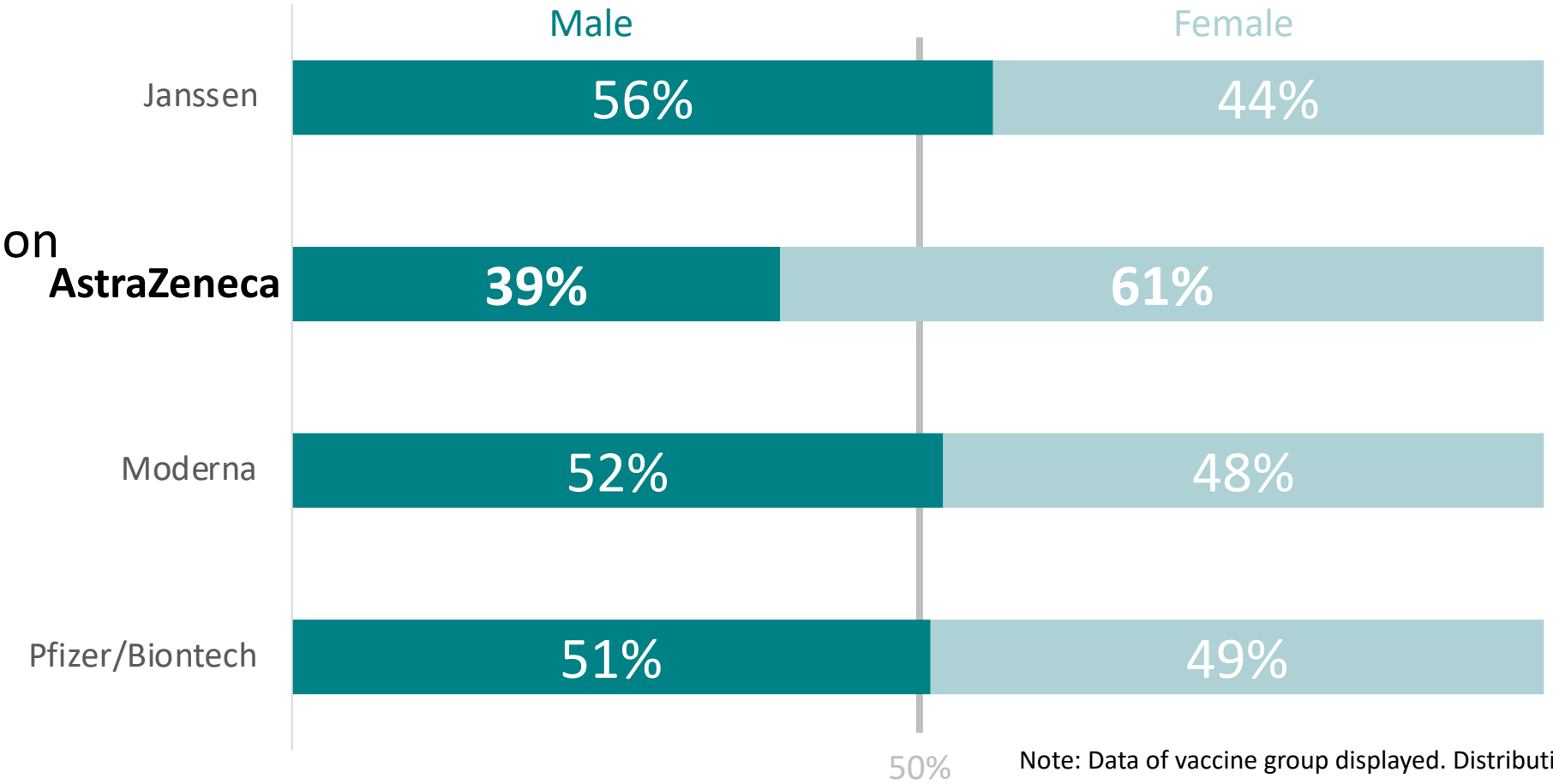
- ✓ Color
- ✓ Fonts
- ✓ Style
- ✓ All Data?
- ✓ Comparison
- ○ Annotate



SFX: Focus Attention

- ✓ Color
- ✓ Fonts
- ✓ Style
- ✓ All Data?
- ✓ Comparison
- ✓ Annotate

Gender Distribution in COVID-19 Vaccine studies

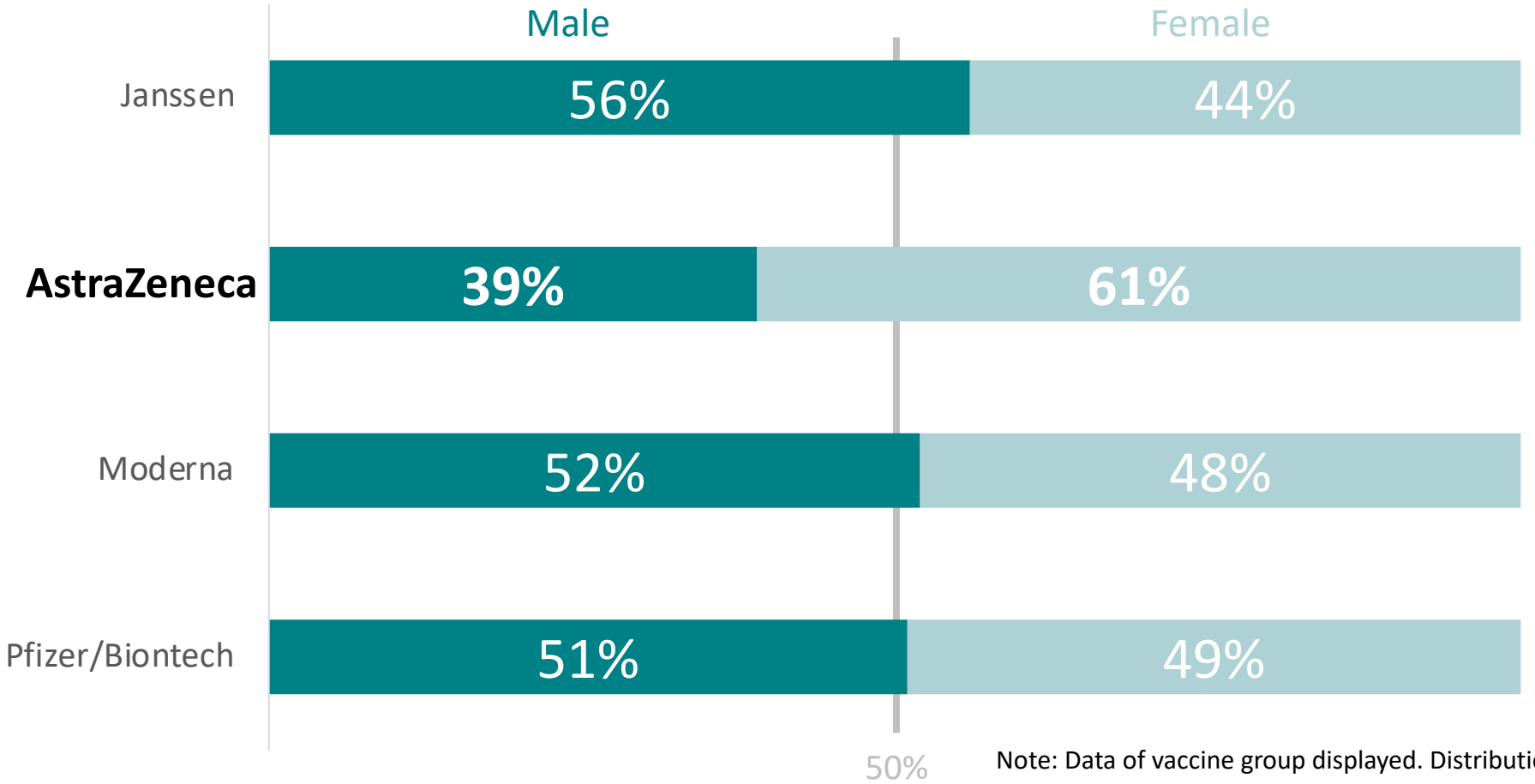


Noticeable differences observed compared to the expected 50:50 gender distribution

Note: Data of vaccine group displayed. Distribution in Placebo group similar.

SFX: Focus Attention

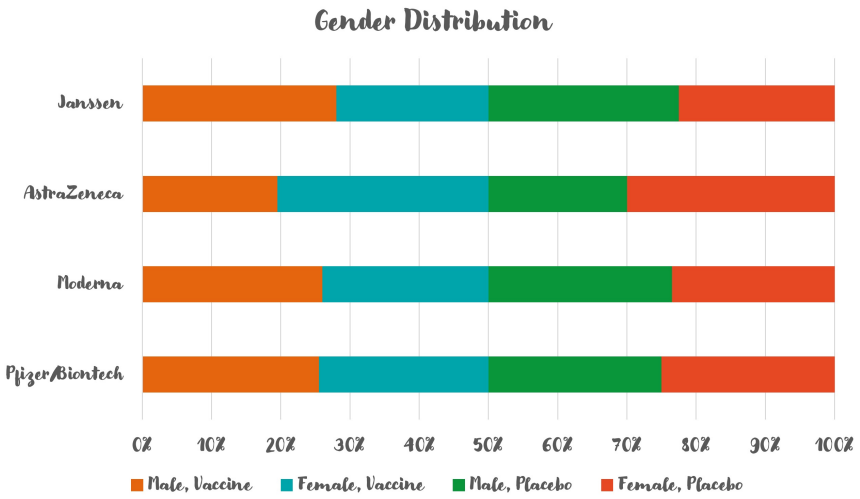
Gender Distribution in COVID-19 Vaccine studies

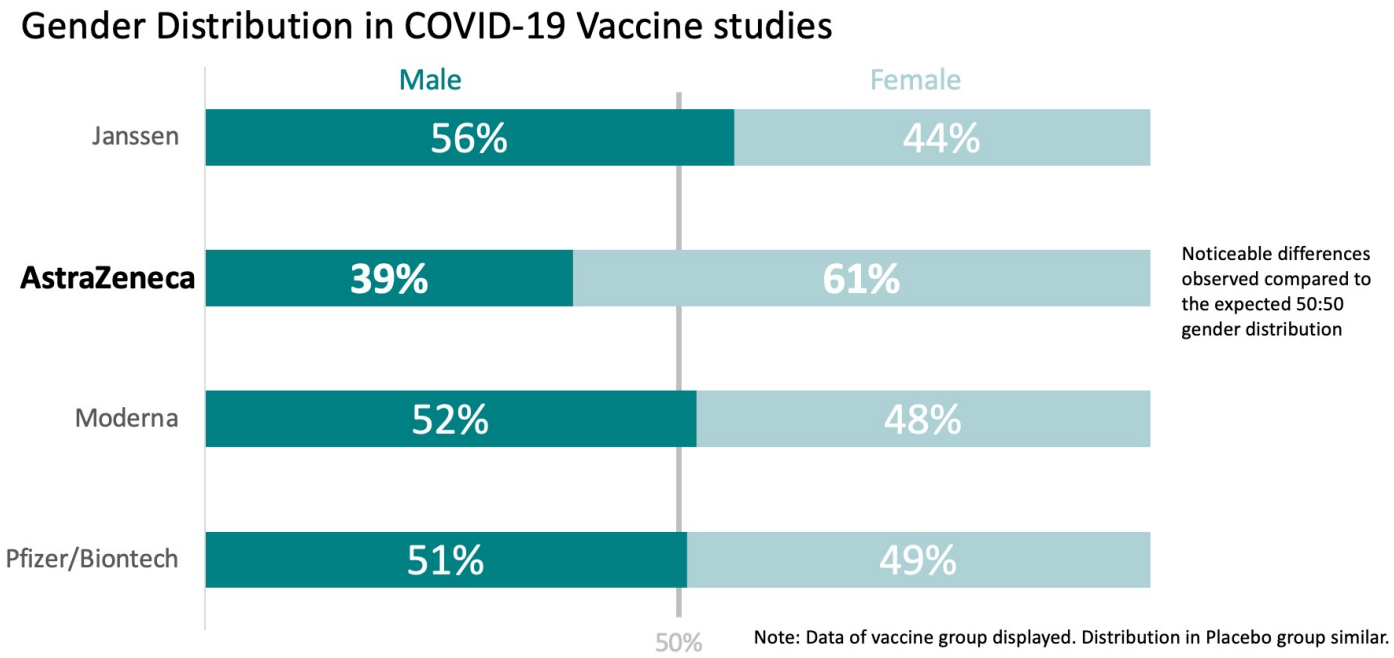


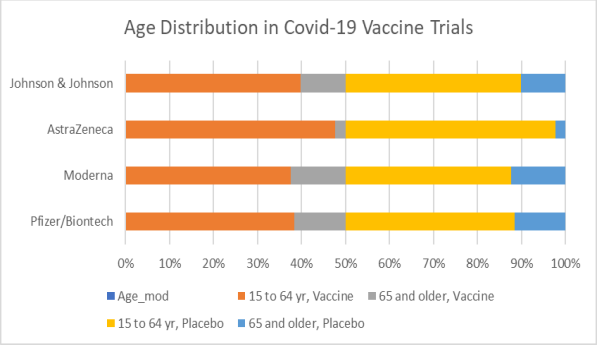
Noticeable differences observed compared to the expected 50:50 gender distribution

Note: Data of vaccine group displayed. Distribution in Placebo group similar.

Special Effects

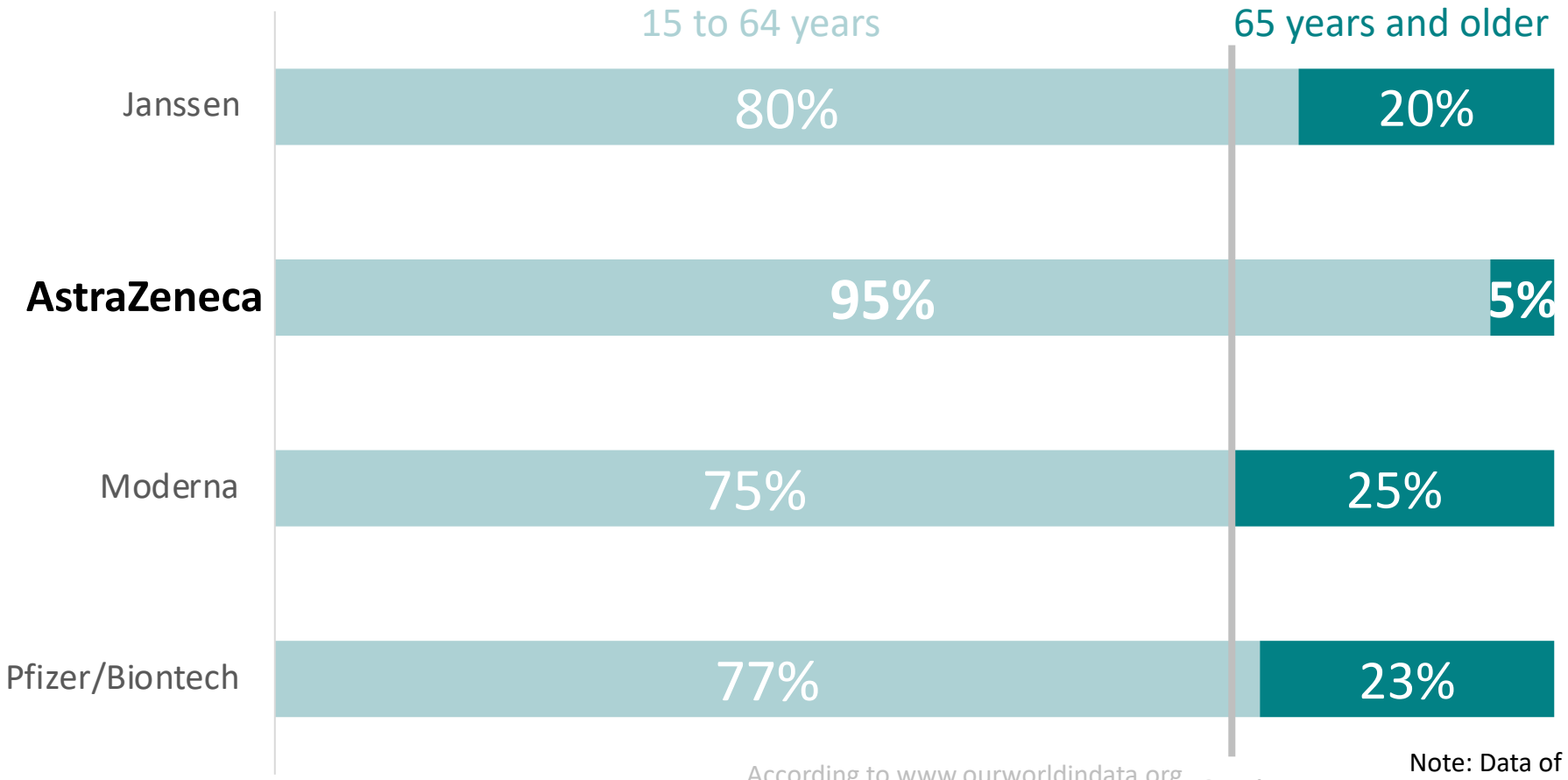






SFX: Age distribution

Age Distribution in COVID-19 Vaccine studies



Noticeable differences observed compared to the expected 75:25 age distribution

According to www.ourworldindata.org
about 25% of a population are in the elderly population

Note: Data of vaccine group displayed. Distribution in Placebo group similar.

What's Next?

- From Data to Insights
- Turn insights into **visuals**
- **Storytelling**
- Storyboarding

Photo by Henry Co on Unsplash

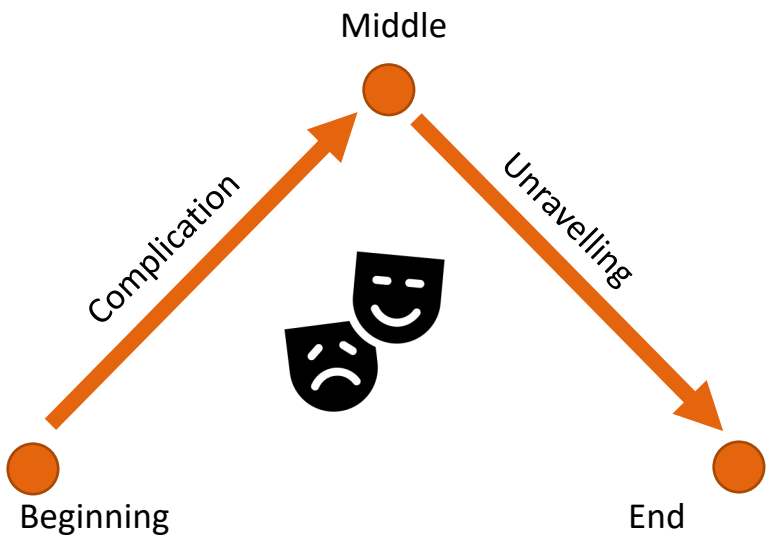
Building the Story

Apply everything we have learned so far and build a powerful study

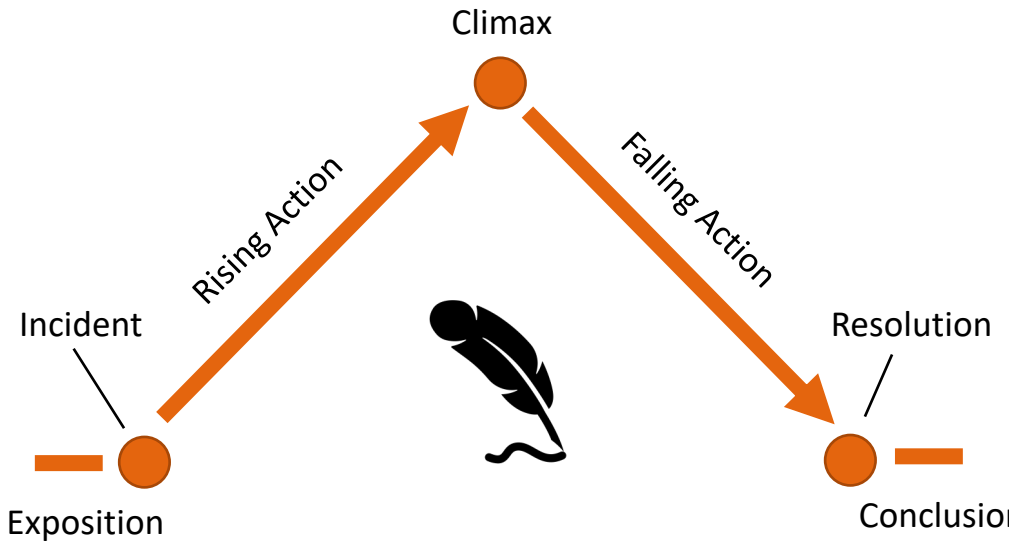
Photo by Leon on Unsplash

Examples for Storytelling Structures

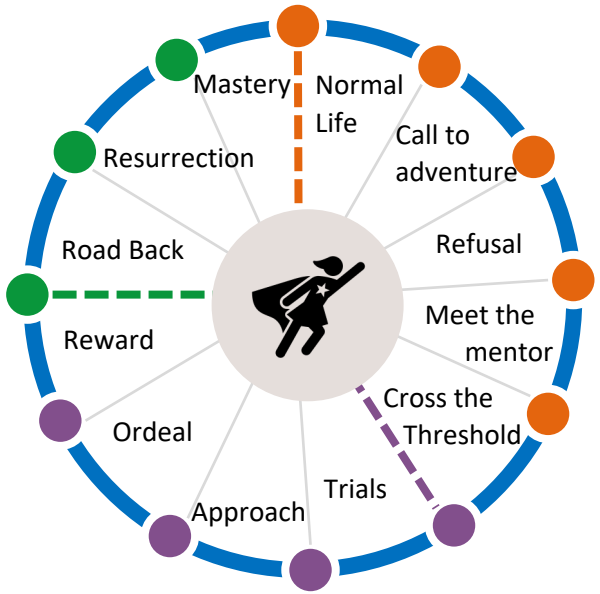
Aristotle's Tragedy Structure



Freytag's Pyramid



Campbells's Hero's Journey



Harry Potter

Star Wars – A new Hope

Data Storytelling Arc

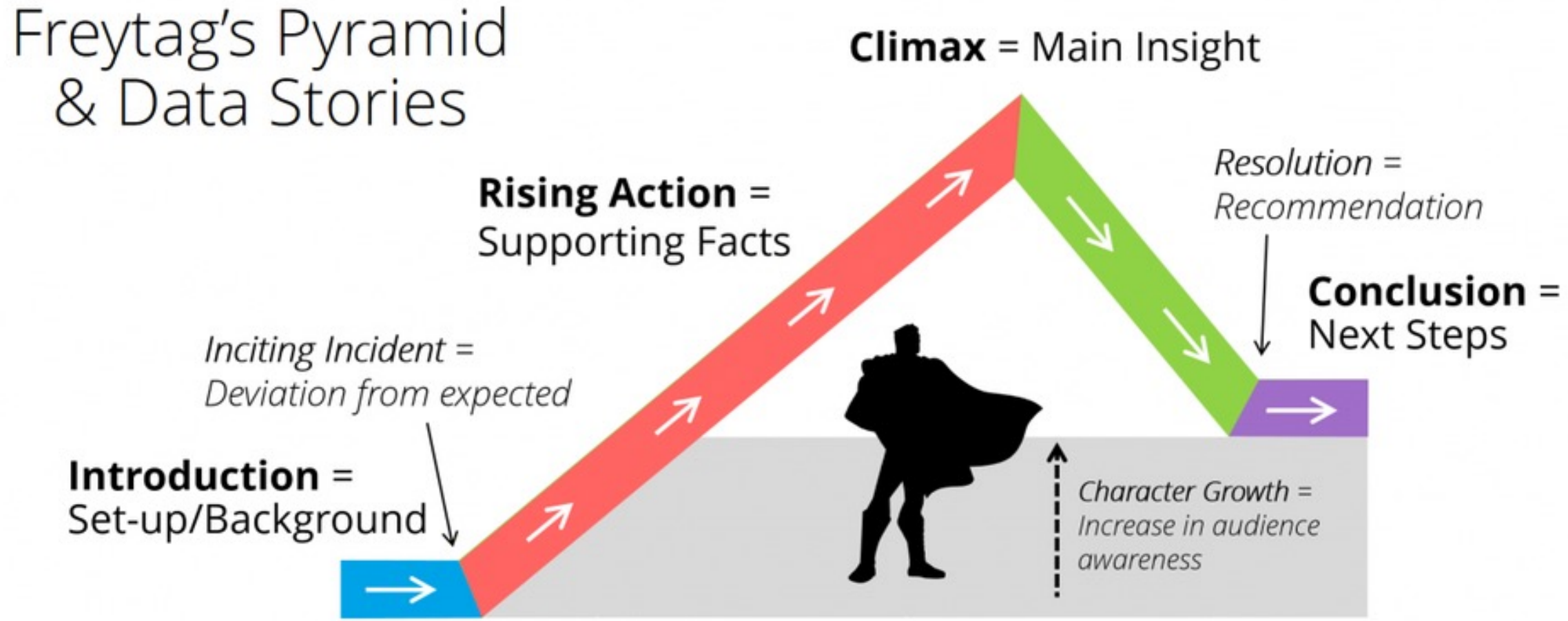


Image Source: https://www.forbes.com/sites/brentdykes/2016/07/13/data-storytelling-separating-fiction-from-facts-2/

Coming Soon

Breakout Room

Storyboarding

Really helpful to use **sticky notes** to actually build your story before you put it into a slide deck

Audience

Study Team to setup a Phase 3 COVID-19 study

Introduction

President Macron publicly stating that AZ vaccine is ineffective

Rising Action

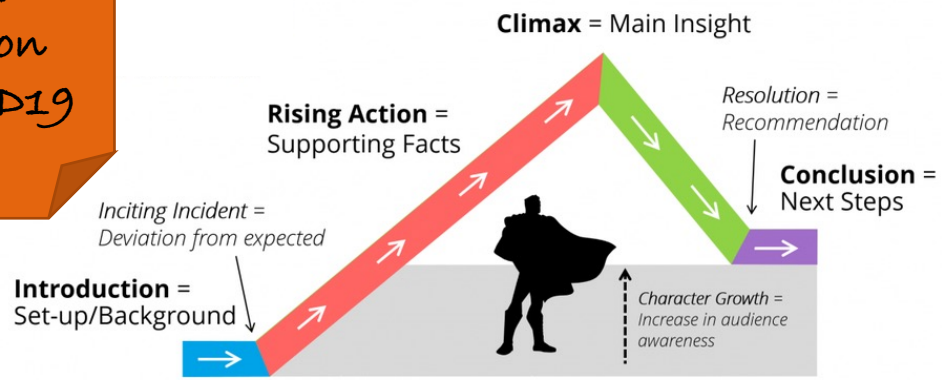
Combining all publicly available information about COVID19 studies

Rising Action

Noticable difference in gender distribution

Climax

Showing our main visual with highlighted differences in age



Conclusion

Recommend to ensure recruiting with anticipated gender and age

BIG Idea

Noticeable difference in AZ vaccine

3-minute story

After the approval of the AZ vaccine, president Macro publicly stated that the AZ vaccine is less effective in elderly patients. We don't want this to happen to our study, do we? It is of critical importance for us to understand how …

Image Source: https://www.forbes.com/sites/brentdykes/2016/07/13/data-storytelling-separating-fiction-from-facts-2/

Coming Soon

Breakout Room

Storyboarding

How could we build this story for your **line manager** in the Data Science department?

Audience

Data Science manager

Introduction

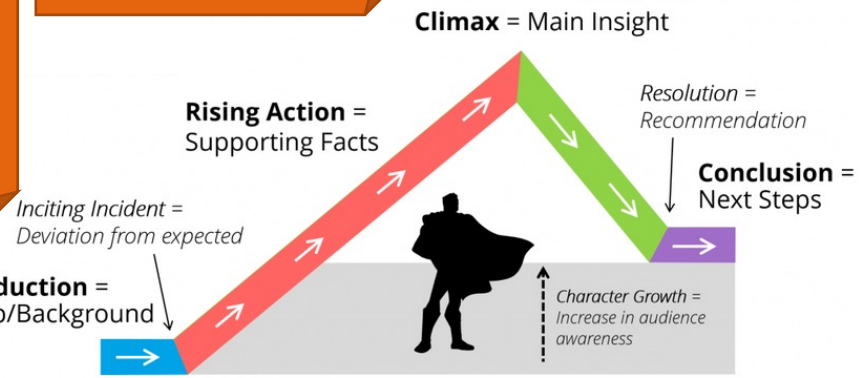
You asked me to find out why president Macron publicly stated that AZ vaccine...

Rising Action

There is a lot of available information about COVID19 studies. Some need transformation ...

Rising Action

Noticable difference in gender distribution



Climax

But what I found then after I transformed everything was a true shock

Conclusion

I need your confirmation that we can present this to the study team to prevent this happens to us

BIG Idea

Fundamental difference in AZ vaccine

3-minute story

After the approval of the AZ vaccine, president Macro publicly stated that the AZ vaccine is less effective in elderly patients. We don't want this to happen to our study, do we? It is of critical importance for us to understand how ...

Image Source: https://www.forbes.com/sites/brentdykes/2016/07/13/data-storytelling-separating-fiction-from-facts-2/

Breakout Rooms

How could we build this story for our **peers** in the Data Science department?

Break out in the same rooms as before.

Discuss how the storytelling elements need to adapt:

- Introduction
- Rising Action
- Climax
- Conclusion

You will have **10 minutes** to discuss

K

Storyboarding

How could we build this story for our **peers** in the Data Science department?

Audience

Data Science peers

Introduction

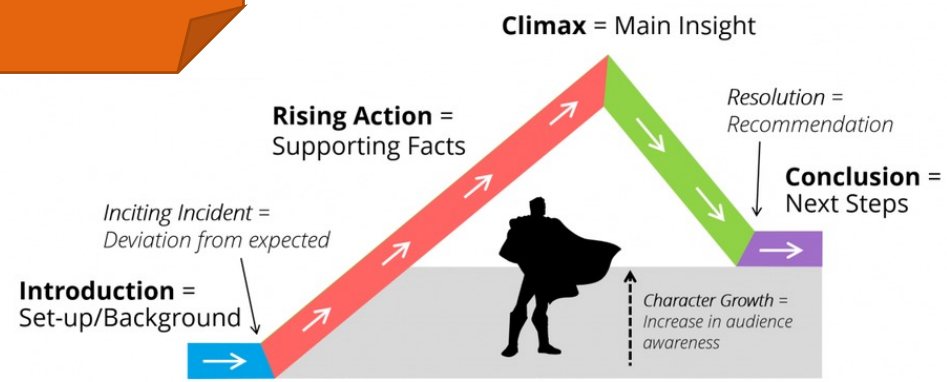
My task was to find out why president Macron publicly stated that AZ vaccine...

Rising Action

There is a lot of available information about COVID19 studies

Climax

At a first glance the data looks comparable, but when you transform everything...



Conclusion

Be careful about your data sources and make them comparable

BIG Idea

Fundamental difference in AZ vaccine

3-minute story

After the approval of the AZ vaccine, president Macro publicly stated that the AZ vaccine is less effective in elderly patients. We don't want this to happen to our study, do we? It is of critical importance for us to understand how ...

Image Source: https://www.forbes.com/sites/brentdykes/2016/07/13/data-storytelling-separating-fiction-from-facts-2/

Learning

- The data is the hero of your story
- Choose an effective visual
- Reduce clutter and ensure focus
- **Take your time to build an effective study**

The END

- Data Storytelling is an art
- Be clear about the **BIG IDEA**
- Build a 3 minute story to be independent from slide decks
- Know your Audience
- Build powerful data **visuals to tell your story**
- Build your storyboard to get your data insights across

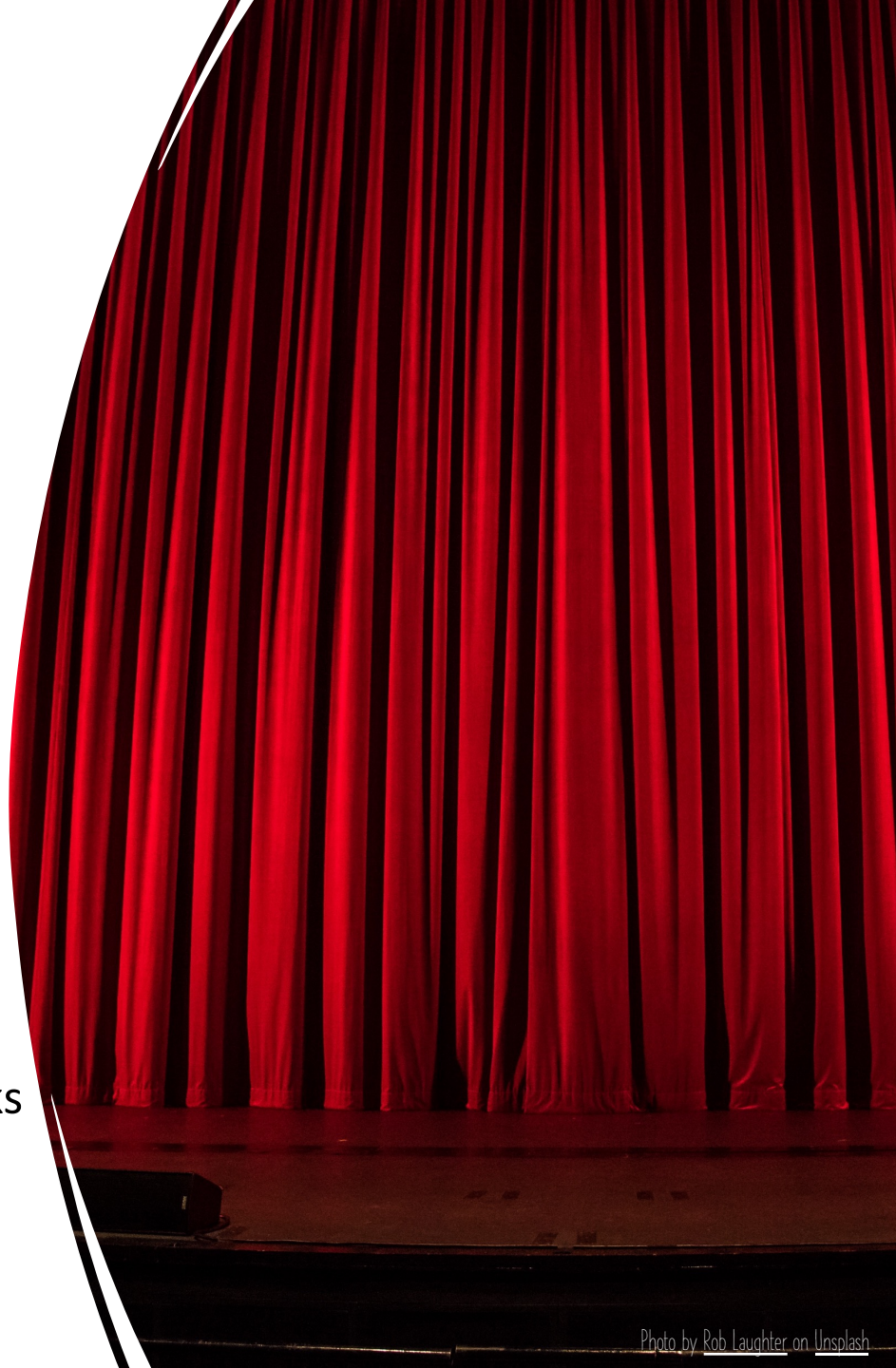


Photo by Rob Laughter on Unsplash

Recommended Readings