

# ARTICLE VIRALITY

Olivia Encarcion  
Phillip Lin  
Leo Yang  
Allison Yang



# SUBJECT INTRODUCTION

**Article Virality** -> What factors make an article go viral?

**Motivation** -

- What makes one online news article popular versus another?
- Popularity = # of shares
- Rise of internet
- What makes something trendy

**Research Question:** How do different article attributes (ex. Polarity, Positive/Negative Sentiments, Number of Images, etc.) relate to its virality on social media?



# DATA SET INTRODUCTION

Source: University of California Irvine Machine Learning Repository

Title: “Online News Popularity”

Source of Articles: Mashable (2013-2015)

Overview:

- Objective: Predict the number of shares/virality of articles based on various features.
- Size: 39,644 observations (articles)
- Time Frame: 2013-2015

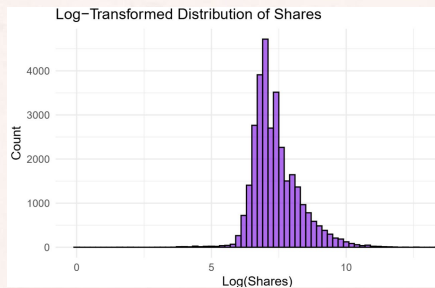
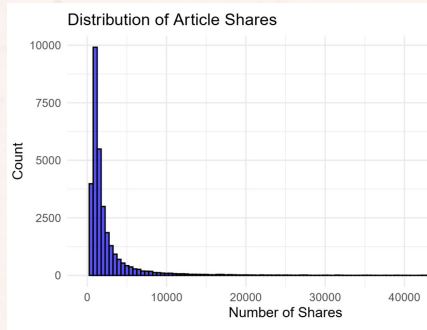
Key Variables:

- Shares (Target): Number of shares for each article
- Rate of Positive Words: Percentage of positive words in non-neutral tokens (words that convey emotion)
- Rate of Negative Words: Percentage of negative words in non-neutral tokens (words that convey emotion)
- Title Sentiment Polarity: Measure of the title’s emotional polarity
- N Tokens in Content: Number of words in the content of the article
- N Tokens in Title: Number of words in the article’s title
- Day Published: Merged categorical variable representing the day of the week
- Data Channel: Merged categorical variable representing article topics (e.g., entertainment, business, tech, etc.)

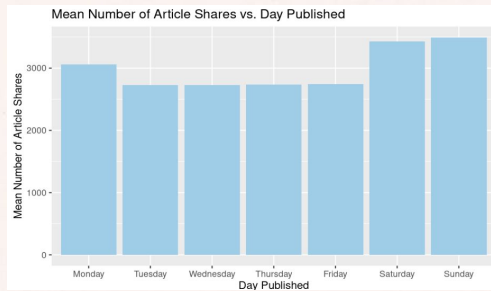
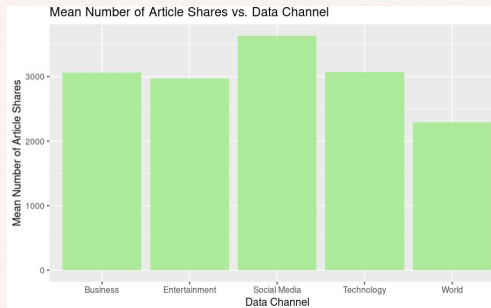
**Response = Shares**

Highly right skewed →

Log transform shares

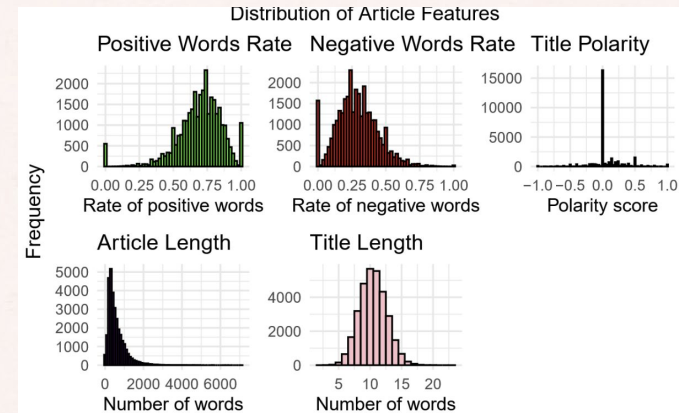
**Median**: 1400**Mean**: 2878**Categorical  
(Day pub./Data Channel)**

- weekends performed better
- Social media performed best but also had least amount of entries

**Univariate Dist.**

Pos. Rate, Neg. rate, Title  
Polarity, Article Length, Title  
Length dists.

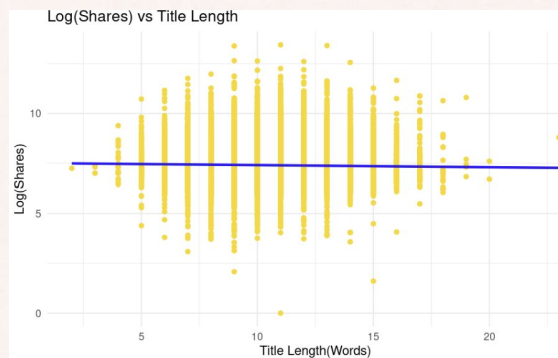
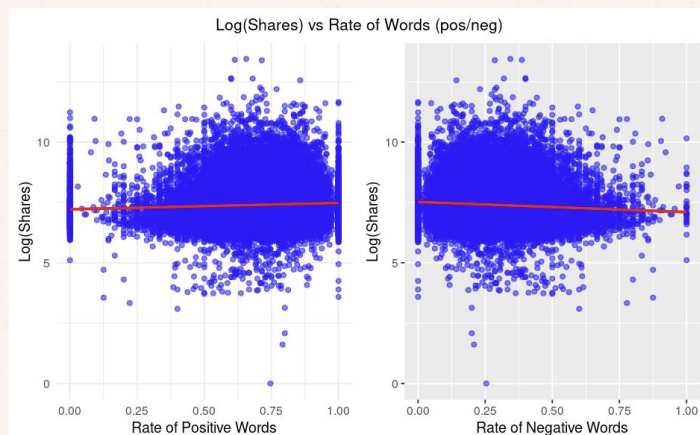
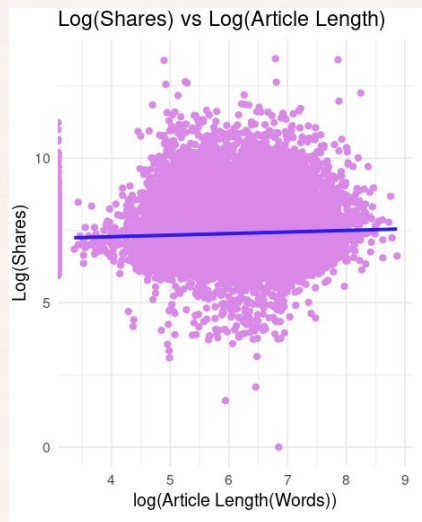
- Strong right skew for article length
- Extreme mode for title polarity





## Log(shares) vs. Log(article length), title length, rate of words

Plotting these predictors against log(shares) didn't yield a strong correlation for any of these variables.



## Log(Shares) vs. Data Channel / Title Sentiment / Title Length

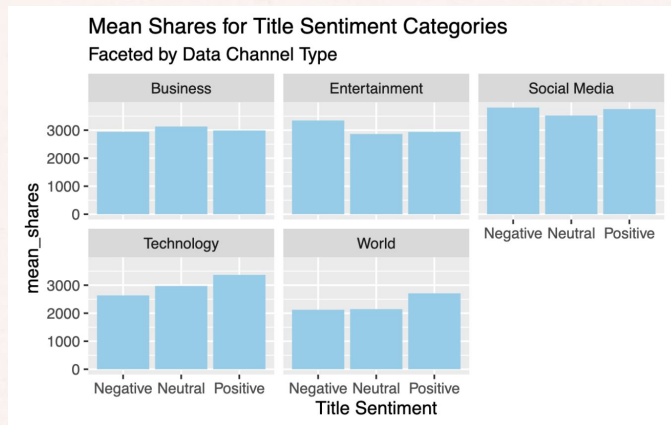
- Initial visualization comparing mean shares and title sentiment suggested possible interaction effects

### Log(Shares) vs. Data Channel + Title Sentiment

- We found that none of the interaction terms were statistically significant though main effect were

### Log(Shares) vs. Data Channel + Title Length

- Title Length and Data Channel World interaction term was statistically significant
- Some of the categorical indicator predictors were not significant, suggesting weak main effect relationship



| term                                     | estimate | std.error | statistic | p.value |
|--|----------|-----------|-----------|---------|
| (Intercept)                              | 7.430    | 0.052     | 141.723   | 0.000   |
| n_tokens_title                           | -0.002   | 0.005     | -0.417    | 0.677   |
| data_channelEntertainment                | -0.150   | 0.076     | -1.978    | 0.048   |
| data_channelSocial Media                 | 0.419    | 0.098     | 4.270     | 0.000   |
| data_channelTechnology                   | 0.292    | 0.072     | 4.064     | 0.000   |
| data_channelWorld                        | -0.415   | 0.071     | -5.827    | 0.000   |
| n_tokens_title:data_channelEntertainment | 0.005    | 0.007     | 0.676     | 0.499   |
| n_tokens_title:data_channelSocial Media  | -0.005   | 0.010     | -0.554    | 0.579   |
| n_tokens_title:data_channelTechnology    | -0.012   | 0.007     | -1.712    | 0.087   |
| n_tokens_title:data_channelWorld         | 0.020    | 0.007     | 2.956     | 0.003   |

## NEXT STEPS

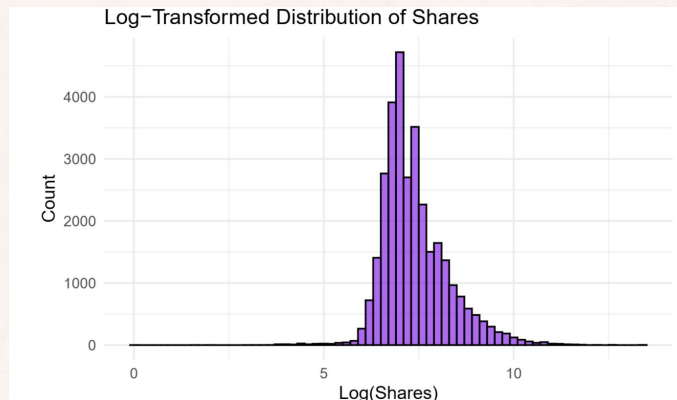
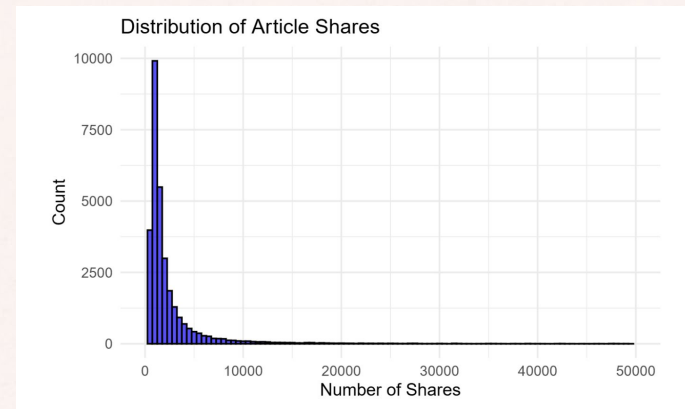
- Exploring additional variables to find more statistically significant predictors
- Experiment with breaking pos/neg word rate into bins to investigate possible simpsons effect
- Construct a MLR model w/ a bottom-up approach
  - Perhaps trying Day Published + Data Channel first as based on EDA seem most important

-

## RESPONSE VARIABLE

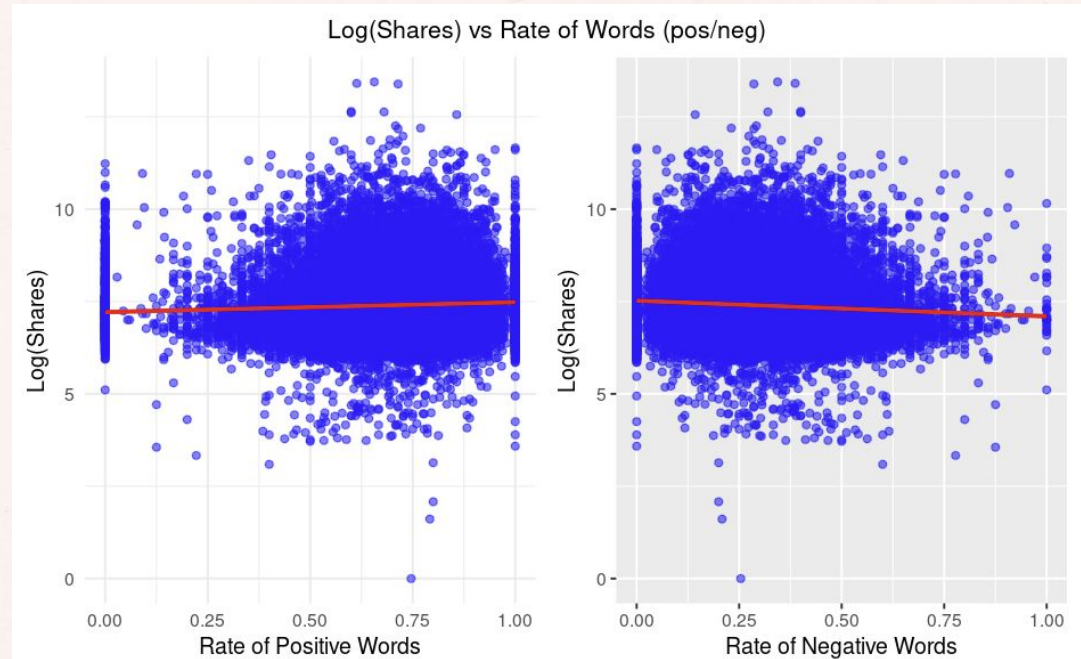
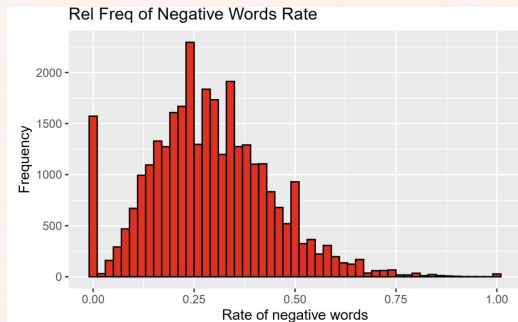
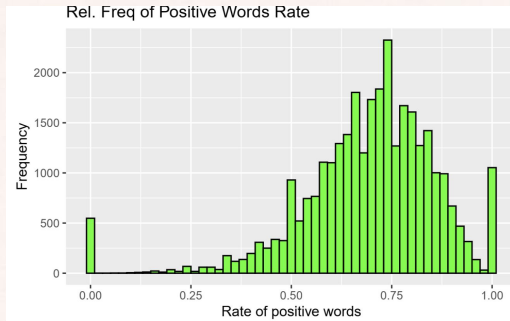
The dist. of shares is highly right skewed, with a median of only 1400 shares vs a far higher mean of 2878.

To reduce the effect of influential outliers and make interpretation easier, we will log-transform shares.



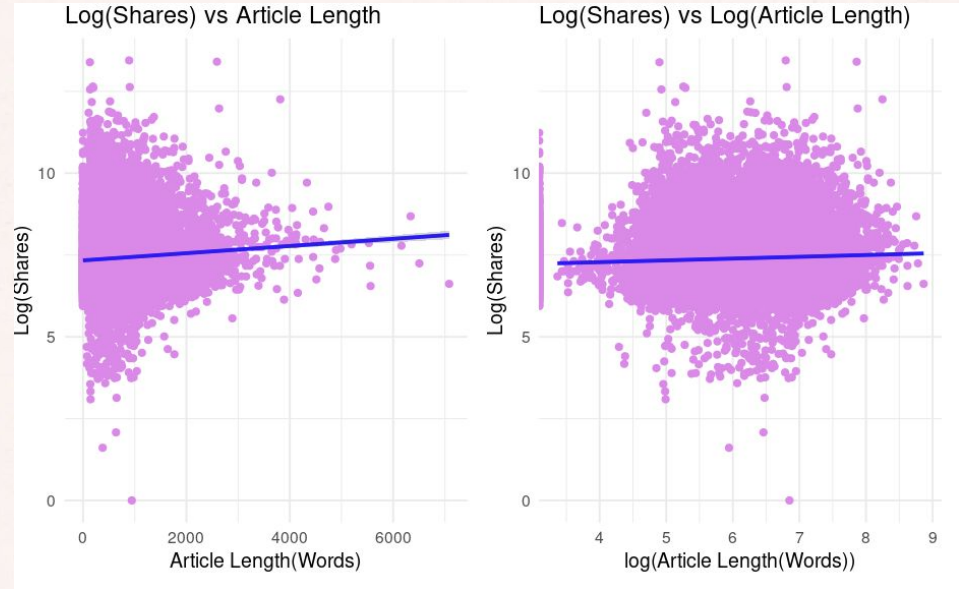
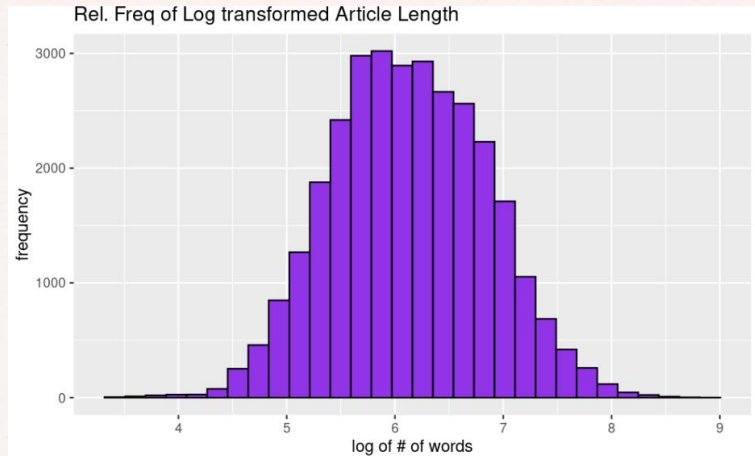


# POSITIVE & NEGATIVE WORD RATES

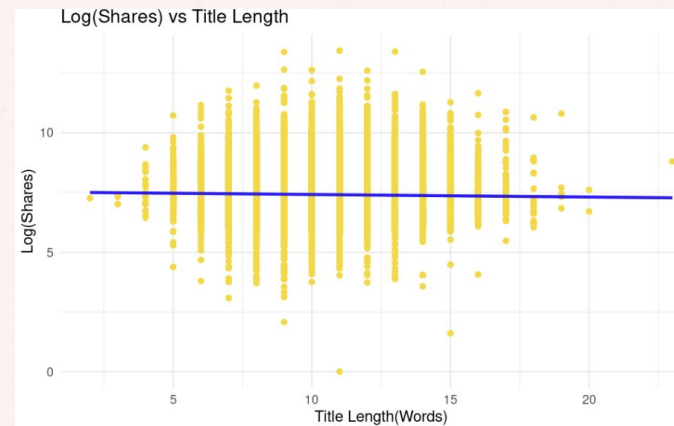
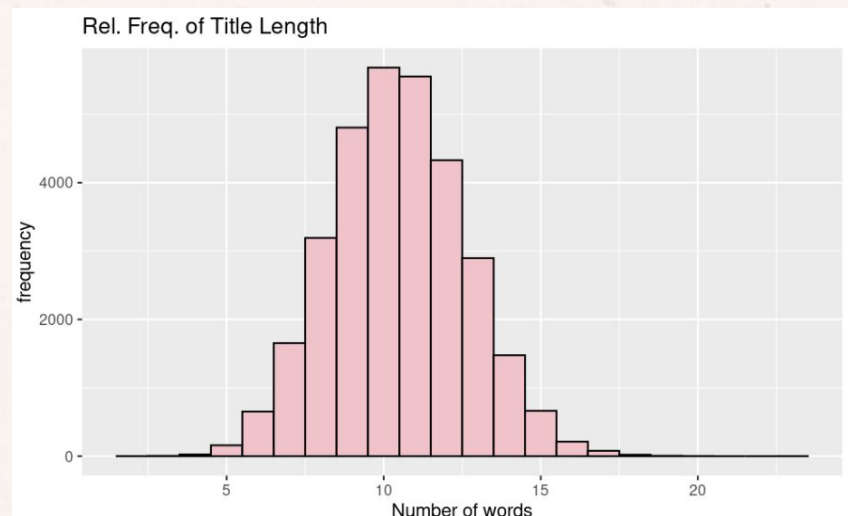


```
[1] "Variance Inflation Factors:"
rate_negative_words rate_positive_words
1.946291           1.946291
```

# ARTICLE LENGTH

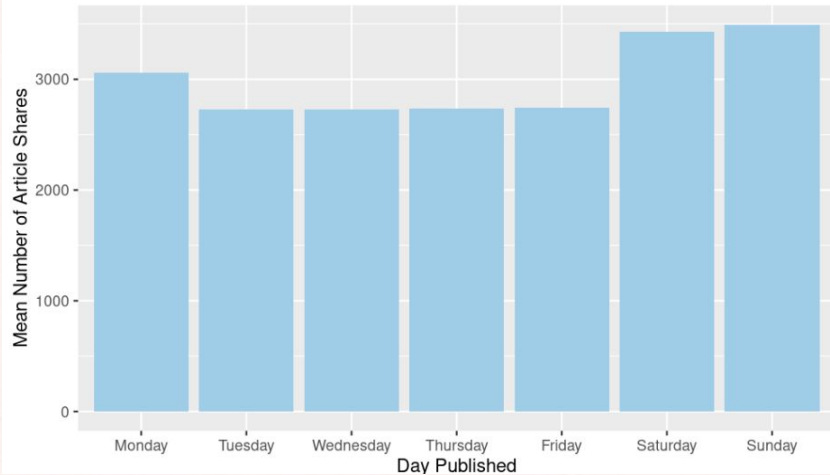


# TITLE LENGTH

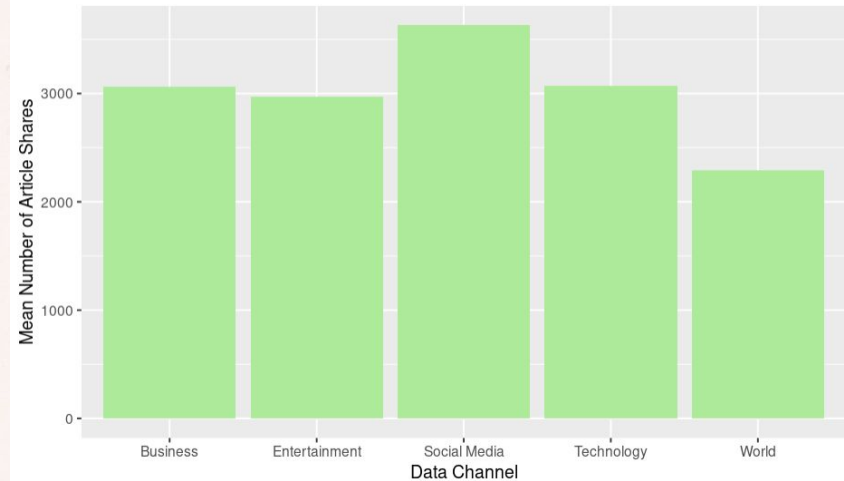


# CATEGORICAL VARIABLES

Mean Number of Article Shares vs. Day Published



Mean Number of Article Shares vs. Data Channel



## INTERACTION EFFECTS - LOG(SHARES)

| term                                     | estimate | std.error | statistic | p.value |
|--|----------|-----------|-----------|---------|
| (Intercept)                              | 7.430    | 0.052     | 141.723   | 0.000   |
| n_tokens_title                           | -0.002   | 0.005     | -0.417    | 0.677   |
| data_channelEntertainment                | -0.150   | 0.076     | -1.978    | 0.048   |
| data_channelSocial Media                 | 0.419    | 0.098     | 4.270     | 0.000   |
| data_channelTechnology                   | 0.292    | 0.072     | 4.064     | 0.000   |
| data_channelWorld                        | -0.415   | 0.071     | -5.827    | 0.000   |
| n_tokens_title:data_channelEntertainment | 0.005    | 0.007     | 0.676     | 0.499   |
| n_tokens_title:data_channelSocial Media  | -0.005   | 0.010     | -0.554    | 0.579   |
| n_tokens_title:data_channelTechnology    | -0.012   | 0.007     | -1.712    | 0.087   |
| n_tokens_title:data_channelWorld         | 0.020    | 0.007     | 2.956     | 0.003   |



# TABLE OF CONTENTS

01

## OUR TEACHING METHOD

You can describe the topic of the section here

02

## OUR ACADEMIC AREAS

You can describe the topic of the section here

03

## OUR TEACHERS

You can describe the topic of the section here

04

## ENROLLMENT PROCESS

You can describe the topic of the section here

01

# OUR TEACHING METHOD

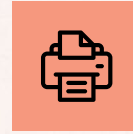
You can enter a subtitle here if you need it

# MAYBE YOU NEED TO DIVIDE THE CONTENT



## MERCURY

It's the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than the Moon



## VENUS

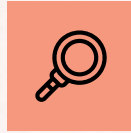
Venus has a beautiful name and is the second planet from the Sun. It's hot and has a poisonous atmosphere

# YOU CAN USE THREE COLUMNS, WHY NOT?



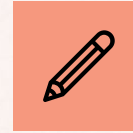
## MERCURY

It's the closest planet to the Sun and the smallest in the Solar System



## MARS

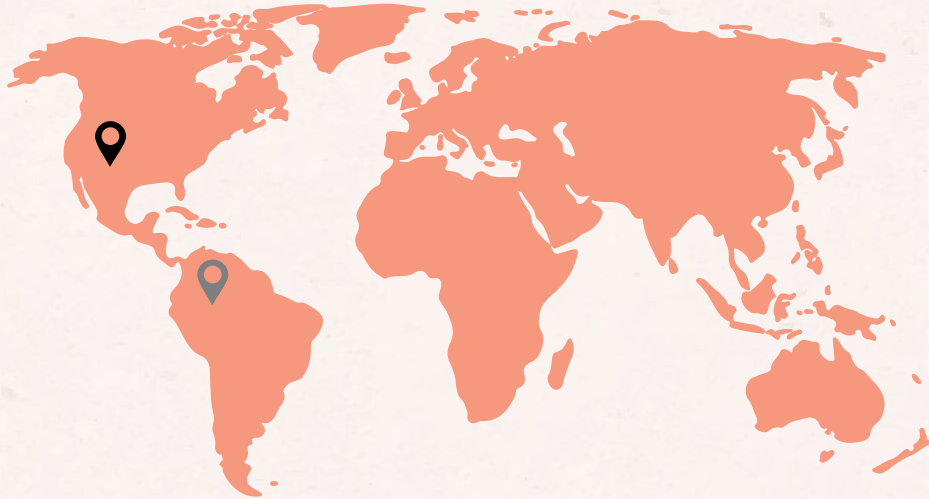
Despite being red, Mars is actually a cold place. It's full of iron oxide dust



## VENUS

Venus has a beautiful name and is the second planet from the Sun

# THIS IS A MAP



## VENUS



Venus is the second planet  
from the Sun

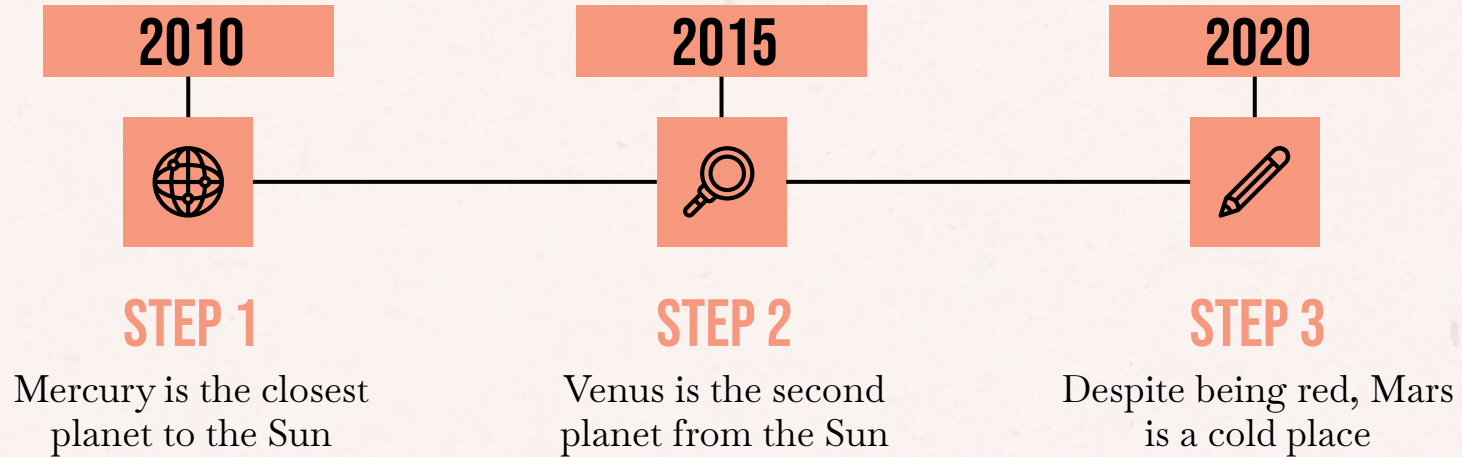
## MERCURY



Mercury is the closest  
planet to the Sun



# INFOGRAPHICS MAKE YOUR IDEA UNDERSTANDABLE...



## ...AND THE SAME GOES FOR TABLES

|                | TEST 01                                 | TEST 02                                   | TEST 03                                |
|----------------|---|---|--|
| 1ST EVALUATION | Pluto is considered a dwarf planet      | Jupiter is the biggest planet of them all | Mercury is the smallest planet         |
| 2ND EVALUATION | Venus is the second planet from the Sun | Mars is actually a very cold place        | Earth is the third planet from the Sun |
| 3RD EVALUATION | Ceres is in the main asteroid belt      | The Sun is the star we orbit around       | Neptune is far away from Earth         |

# THANKS!

Do you have any questions?

youremail@freepik.com  
+91 620 421 838  
yourcompany.com



Please keep this slide for attribution

CREDITS: This presentation template was created  
by **Slidesgo**, including icons by **Flaticon** and  
infographics & images by **Freepik**



# RESOURCES

Did you like the resources on this template? Get them for free at our other websites:

## PHOTOS:

- Woman studying on laptop
- Close up blurry students at university
- University colleagues talking in the library
- Businessman at office with laptop
- Young woman working with her headphones

# Instructions for use

If you have a free account, in order to use this template, you must credit [Slidesgo](#) by keeping the [Thanks](#) slide. Please refer to the next slide to read the instructions for premium users.

## **As a Free user, you are allowed to:**

- Modify this template.
- Use it for both personal and commercial projects.

## **You are not allowed to:**

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>



# Instructions for use (premium users)

As a Premium user, you can use this template without attributing Slidesgo or keeping the "Thanks" slide.

## You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the "Thanks" slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

## You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

# Fonts & colors used

This presentation has been made using the following fonts:

## **Bebas Neue**

(<https://fonts.google.com/specimen/Bebas+Neue>)

## **Baskerville**

(<https://fonts.google.com/specimen/Baskerville>)

#f6987e

#000000

#efe7e5

#7f7f7f

#a5a5a5

#bfbfbf

# Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out [How it Works](#).



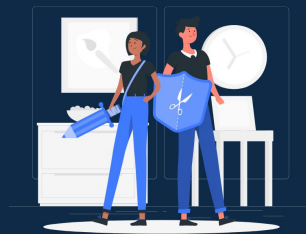
Pana



Amico



Bro



Rafiki



Cuate

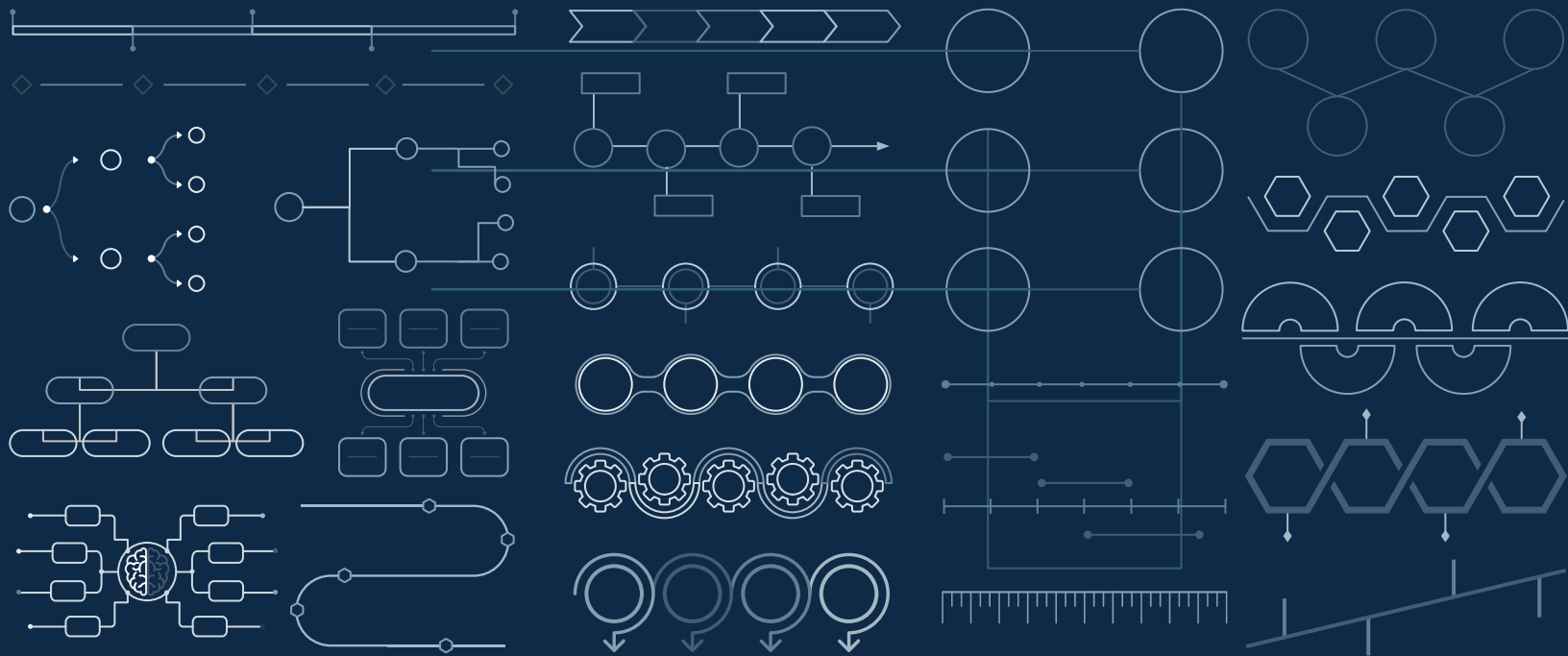
# Use our editable graphic resources...

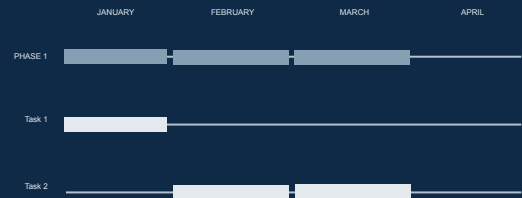
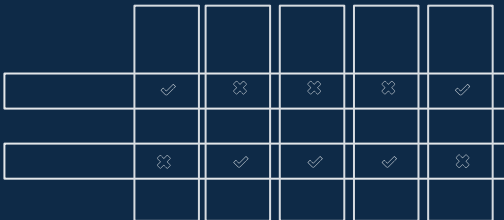
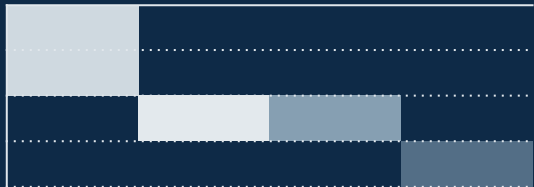
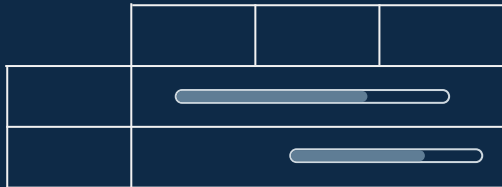
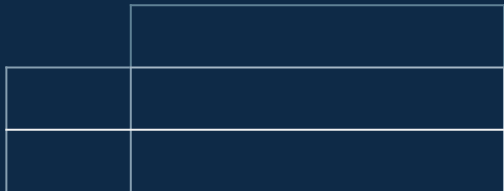
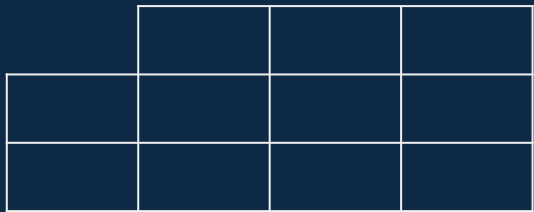
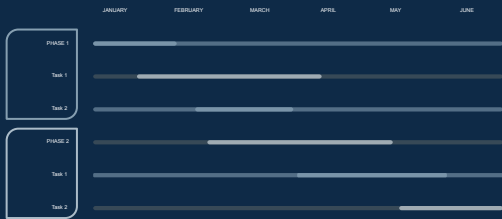
You can easily [resize](#) these resources without losing quality. To [change the color](#), just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more [infographics](#) on [Slidesgo](#).















# ...and our sets of editable icons

You can resize these icons without losing quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.



## Educational Icons



## Medical Icons





## Business Icons



## Teamwork Icons



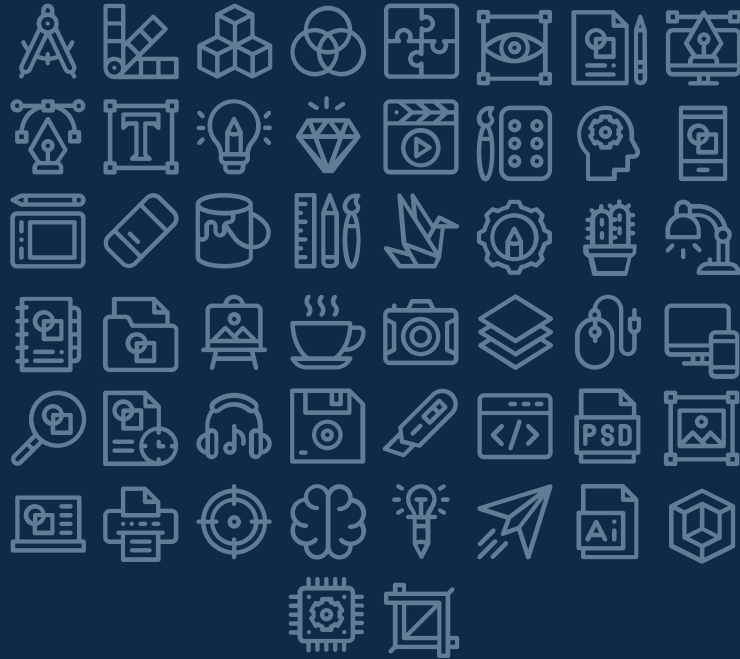
## Help & Support Icons



## Avatar Icons



## Creative Process Icons



## Performing Arts Icons



# Nature Icons



# SEO & Marketing Icons



