

# Exercise 10 | Color and Accessibility

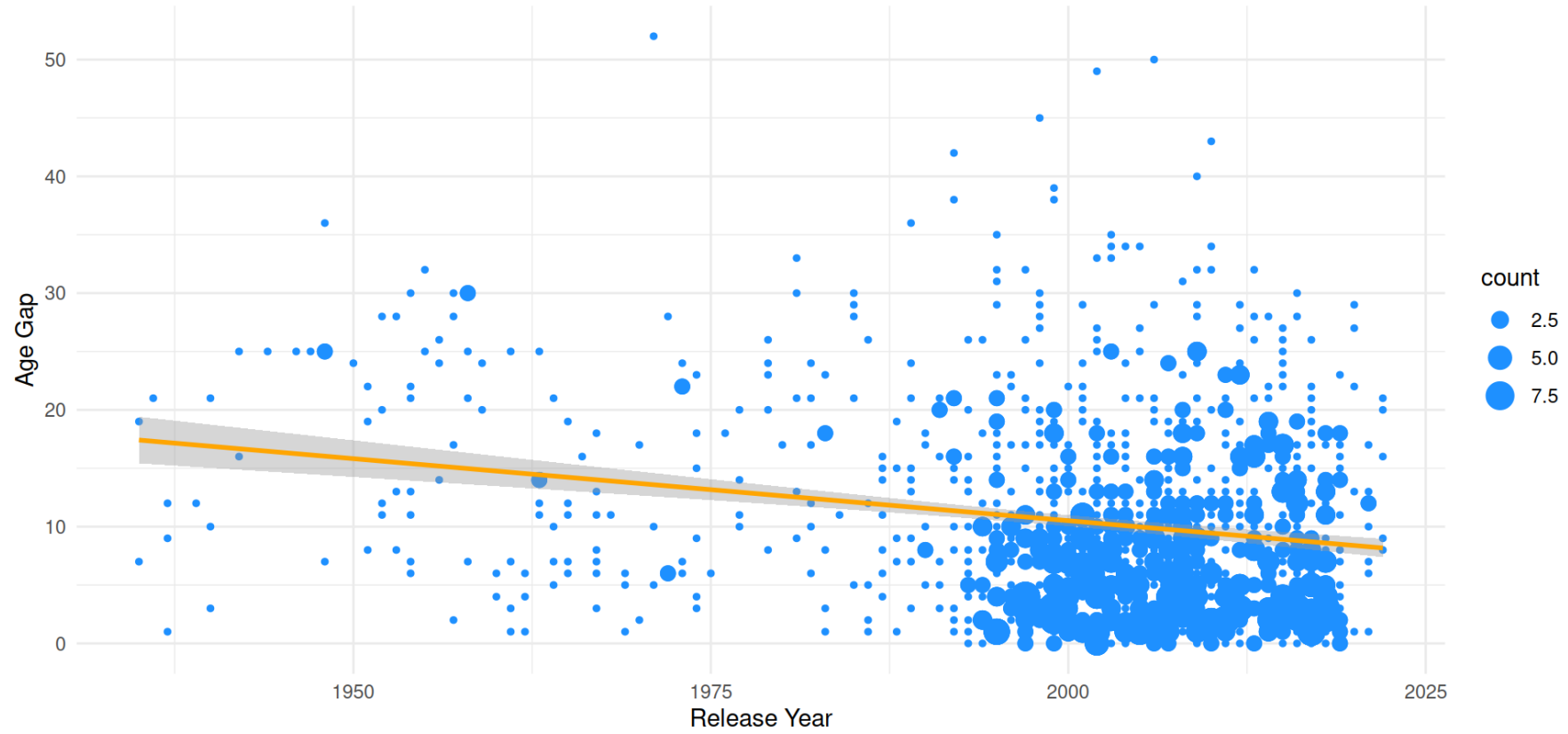
Max Pellert (<https://mpellert.at>)

IS 616: Large Scale Data Analysis and Visualization

**Let's check out some example  
code from the first hand-in  
assignment**

```
1 library("tidyverse")
2 library("scales")
3 library("gridExtra")
4 library("fBasics")
5 library("xtable")
6
7 age_gaps <- read.csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-01-01/age_gaps.csv")
8
9 movies <- read.csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-01-01/movies.csv")
10
11 bechdel_ratings <- read.csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-01-01/bechdel_ratings.csv")
12
13 # Create a count of observations for each pair of release_year and age_difference values
14 age_gaps_count <- age_gaps %>%
15   count(release_year, age_difference, name = "count") %>%
16   ungroup() # not perfectly sure about this one actually, but it works :D
17
18 # I think it's used to make sure every single observation is assigned a value
19 # Add the count variable to the age_gaps dataset
20 age_gaps <- age_gaps %>%
21   left_join(age_gaps_count, by = c("release_year", "age_difference"))
22
23 # Plot age gaps against release years
```

Age Gaps in Romantic Movies by Year

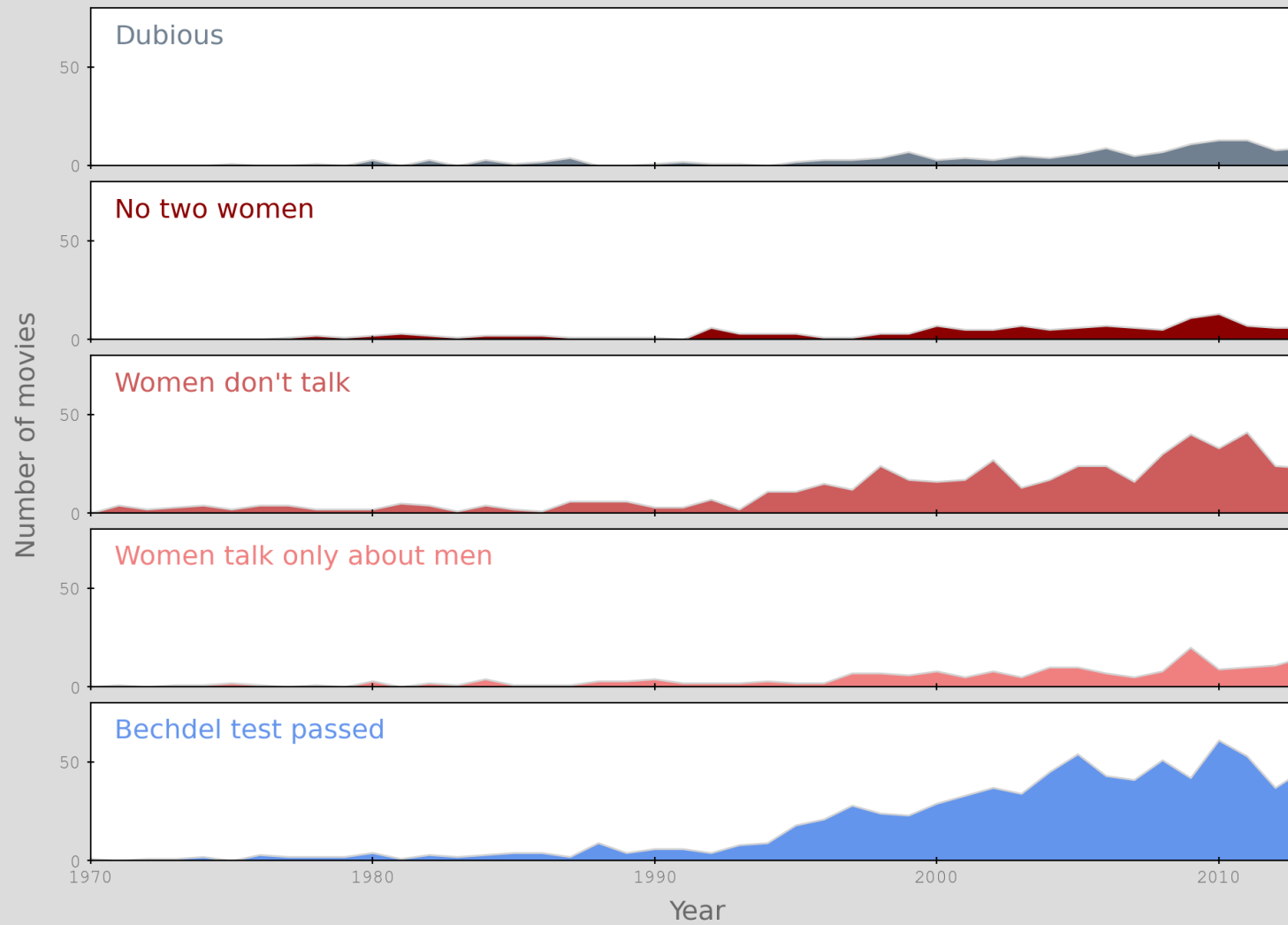


```
1 # import libraries
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import numpy as np
5
6 # load movies table
7 movies = pd.read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data')
8 # select relevant columns
9 movies = movies[['year', 'clean_test']]
10 # rename test column
11 movies = movies.rename(columns={'clean_test': 'bechdel_test'})
12 # movies.head()
13
14 # bins that represent the years = x axis
15 bins = np.arange(1969, 2014, 1)
16
17 # Bechdel classes and their label and color used in graphic
18 bechdel_list = [
19 ('dubious', 'Dubious', 'slategray'),
20 ('nowomen', 'No two women', 'darkred'),
21 ('notalk', "Women don't talk", 'indianred'),
22 ('men', 'Women talk only about men', 'lightcoral'),
23 ('ok', 'Bechdel test passed', 'cornflowerblue')
```

```
1 # set background color of figure
2 fig.set_facecolor(color='gainsboro')
3
4 # common axes labels
5 # add a big axis, hide frame
6 fig.add_subplot(111, frameon=False)
7
8 # hide tick and tick label of the big axes
9 plt.tick_params(labelcolor='none', top=False, bottom=False, left=False, right=False)
10
11 # set axes names
12 plt.xlabel('Year', color='dimgray', fontsize='x-large')
13 plt.ylabel('Number of movies', color='dimgray', fontsize='x-large')
14
15 # set axis limits and values for first axis (axes are shared so it influences all axes)
16 axs[0].set(
17 xlim=(1970, 2013), ylim=(0, 80),
18 xticks=range(1970, 2020, 10), yticks=(0, 50)
19 )
20
21 # iterate over all axes and bechdel classes stated before
22 for ax, (bechdel_result, label, color) in zip(axs.flat, bechdel_list):
23
```

# Evolution of representation of women in movies

Comparison of the Bechdel test classes over time



**For the second hand-in  
assignment...**





UN Votes

[https://github.com/rfordatascience/tidytuesday/  
blob/master/data/2021/2021-03-23/readme.md](https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-03-23/readme.md)

# United Nations General Assembly Voting Data

Version 31.0



Voeten, Erik; Strezhnev, Anton; Bailey, Michael, 2009, "United Nations General Assembly Voting Data", <https://doi.org/10.7910/DVN/LEJUQZ>, Harvard Dataverse, V31

[Cite Dataset](#) ▾

Learn about [Data Citation Standards](#).

## Description ?

This is a dataset of roll-call votes in the UN General Assembly 1946-2022 (sessions 1-77). Note that most votes take place in the Fall but some votes also occur in the spring of the following year. Note that for the five most recent years, failed votes and votes on amendments and paragraphs have not yet been added. Moreover, "important votes" are also not yet coded for these years. There will be an update to the raw data, likely in the Fall. Note the codebook as there are some issues with dates and other features of the raw data, which are pieced together from numerous sources. The dataset includes issue codes and descriptions. It also contains ideal point estimates derived from these votes as described in Bailey, Michael A., Anton Strezhnev, and Erik Voeten. "Estimating dynamic state preferences from United Nations voting data." *Journal of Conflict Resolution* 61.2 (2017): 430-456. Code to estimate ideal points and issue ideal points are available at: <https://github.com/evoeten/United-Nations-General-Assembly-Votes-and-Ideal-Points>. The ideal point estimates are based on sessions, not years. There are also dyadic agreement scores, See the Bailey, Strezhnev, Voeten article for why these should not be used in most instances. (2023-07-25)

Voeten, E., Strezhnev, A., & Bailey, M. (2009). United Nations General Assembly Voting Data [dataset]. Harvard Dataverse. <https://doi.org/10.7910/DVN/LEJUQZ>

Voeten, E. (2012). Data and Analyses of Voting in the UN General Assembly. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2111149>

# unvotes.csv

<b>variable</b>	<b>class</b>	<b>description</b>
rcid	double	The roll call id; used to join with un_votes and un_roll_call_issues
country	character	Country name, by official English short name
country_code	character	2-character ISO country code
vote	integer	Vote result as a factor of yes/abstain/no

# roll\_calls.csv

variable	class	description
rcid	integer	.
session	double	Session number. The UN holds one session per year; these started in 1946
importantvote	integer	Whether the vote was classified as important by the U.S. State Department report “Voting Practices in the United Nations”. These classifications began with session 39
date	double	Date of the vote, as a Date vector

# roll\_calls.csv

<b>variable</b>	<b>class</b>	<b>description</b>
unres	character	Resolution code
amend	integer	Whether the vote was on an amendment; coded only until 1985
para	integer	Whether the vote was only on a paragraph and not a resolution; coded only until 1985
short	character	Short description
descr	character	Longer description

# issues.csv

<b>variable</b>	<b>class</b>	<b>description</b>
rcid	integer	The roll call id; used to join with unvotes and un_roll_calls
short_name	character	Two-letter issue codes
issue	integer	Descriptive issue name

# In R

```
1 # Get the Data
2
3 # Read in with tidyuesdayR package
4 # Install from CRAN via: install.packages("tidyuesdayR")
5 # This loads the readme and all the datasets for the week of interest
6
7 # Either ISO-8601 date or year/week works!
8
9 tuesdata <- tidyuesdayR::tt_load('2021-03-23')
10 tuesdata <- tidyuesdayR::tt_load(2021, week = 13)
11
12 unvotes <- tuesdata$unvotes
13
14 # Or read in the data manually
15
16 unvotes <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidyuesday/master')
17 roll_calls <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidyuesday/master')
18 issues <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidyuesday/master')
```

# In Python

```
1 import pandas as pd
2
3 unvotes = pd.read_csv(
4     'https://raw.githubusercontent.com/rfordatascience/'\
5     'tidytuesday/master/data/2021/2021-03-23/unvotes.csv')
6 roll_calls = pd.read_csv(
7     'https://raw.githubusercontent.com/rfordatascience/'\
8     'tidytuesday/master/data/2021/2021-03-23/roll_calls.csv')
9 issues = pd.read_csv(
10    'https://raw.githubusercontent.com/rfordatascience/'\
11    'tidytuesday/master/data/2021/2021-03-23/issues.csv')
```



# Until next time...

Take a look at the United Nations data

Load it, do explorations to see if you run into any issues

After next unit, you will know more about geospatial data visualization and you should use that knowledge with the UN data for your second hand-in assignment