Tidy Tuesday

Eugene Zhou

2023-01-17

TidyTuesday

Join the R4DS Online Learning Community in the weekly #TidyTuesday event! Every week we post a raw dataset, a chart or article related to that dataset, and ask you to explore the data. While the dataset will be "tamed", it will not always be tidy! As such you might need to apply various R for Data Science techniques to wrangle the data into a true tidy format. The goal of TidyTuesday is to apply your R skills, get feedback, explore other's work, and connect with the greater #RStats community! As such we encourage everyone of all skills to participate!

Load the weekly Data

Download the weekly data and make available in the `tt` object.

```
## Only 5 Github queries remaining until 2023-01-21 03:55:03 PM PST.
## Only 5 Github queries remaining until 2023-01-21 03:55:03 PM PST.
## Only 5 Github queries remaining until 2023-01-21 03:55:03 PM PST.
## Only 5 Github queries remaining until 2023-01-21 03:55:03 PM PST.

## Only 5 Github queries remaining until 2023-01-21 03:55:02 PM PST.

## --- Compiling #TidyTuesday Information for 2023-01-17 ----

## Only 4 Github queries remaining until 2023-01-21 03:55:02 PM PST.

## --- There is 1 file available ---

## Only 3 Github queries remaining until 2023-01-21 03:55:02 PM PST.

## --- Starting Download ---

## Only 3 Github queries remaining until 2023-01-21 03:55:02 PM PST.

##  Downloading file 1 of 1: ‘artists.csv‘

## Only 2 Github queries remaining until 2023-01-21 03:55:02 PM PST.

## --- Download complete ---
```

Readme

Take a look at the readme for the weekly data to get insight on the dataset. This includes a data dictionary, source, and a link to an article on the data.

```
tt
```

This dataset is comprised of data related to artists and their work in Janson's History of Art and Gardner's Art Through the Ages.

Glimpse Data

Take an initial look at the format of the data available.

```
tt %>%
  map(glimpse)
```

```
## Rows: 3,162
## Columns: 14
## $ artist_name                <chr> "Aaron Douglas", "Aaron Douglas", "Aaron Do~
## $ edition_number             <dbl> 9, 10, 11, 12, 13, 14, 15, 16, 14, 15, 16, ~
## $ year                       <dbl> 1991, 1996, 2001, 2005, 2009, 2013, 2016, 2~
## $ artist_nationality         <chr> "American", "American", "American", "Americ~
## $ artist_nationality_other   <chr> "American", "American", "American", "Americ~
## $ artist_gender              <chr> "Male", "Male", "Male", "Male", "Male", "Ma~
## $ artist_race                <chr> "Black or African American", "Black or Afri~
## $ artist_ethnicity           <chr> "Not Hispanic or Latino origin", "Not Hispa~
## $ book                       <chr> "Gardner", "Gardner", "Gardner", "Gardner",~
## $ space_ratio_per_page_total <dbl> 0.3533658, 0.3739470, 0.3032593, 0.3770489,~
## $ artist_unique_id           <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 6, 6, 6, 6~
## $ moma_count_to_year         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ whitney_count_to_year      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ artist_race_nwi            <chr> "Non-White", "Non-White", "Non-White", "Non~
```

```
## $artists
## # A tibble: 3,162 x 14
##    artist_~1 editi~2  year artis~3 artis~4 artis~5 artis~6 artis~7 book  space~8
##    <chr>       <dbl> <dbl> <chr>   <chr>   <chr>   <chr>   <chr>   <chr>   <dbl>
##  1 Aaron Do~       9  1991 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.353
##  2 Aaron Do~      10  1996 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.374
##  3 Aaron Do~      11  2001 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.303
##  4 Aaron Do~      12  2005 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.377
##  5 Aaron Do~      13  2009 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.398
##  6 Aaron Do~      14  2013 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.453
##  7 Aaron Do~      15  2016 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.482
##  8 Aaron Do~      16  2020 Americ~ Americ~ Male    Black ~ Not Hi~ Gard~   0.544
##  9 Adélaïde~      14  2013 French  French  Female  White   Not Hi~ Gard~   0.409
## 10 Adélaïde~      15  2016 French  French  Female  White   Not Hi~ Gard~   0.413
## # ... with 3,152 more rows, 4 more variables: artist_unique_id <dbl>,
## #   moma_count_to_year <dbl>, whitney_count_to_year <dbl>,
## #   artist_race_nwi <chr>, and abbreviated variable names 1: artist_name,
```

```
## #   2: edition_number, 3: artist_nationality, 4: artist_nationality_other,
## #   5: artist_gender, 6: artist_race, 7: artist_ethnicity,
## #   8: space_ratio_per_page_total
```

Wrangle

Explore the data and process it into a nice format for plotting! Access each dataset by name by using a dollarsign after the `tt` object and then the name of the data set.

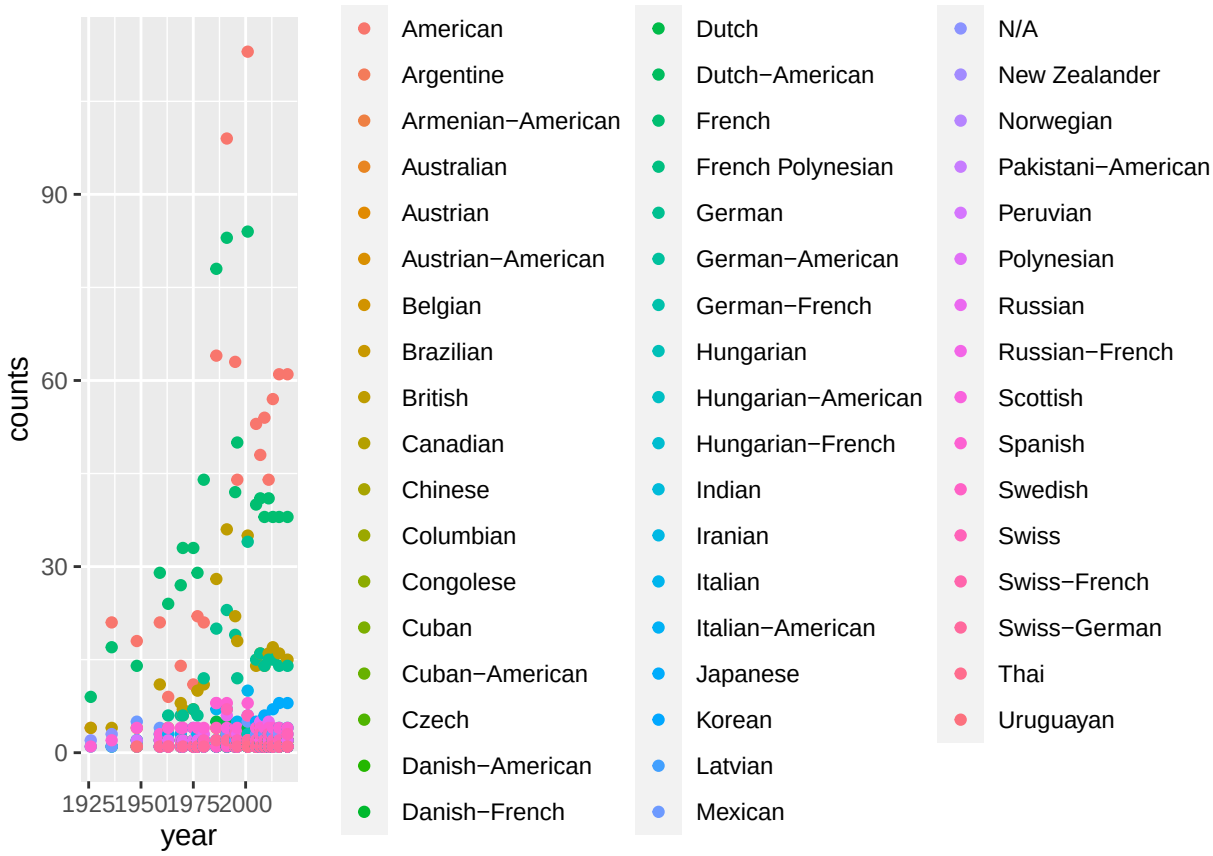
```
# extract art data frame from the tidy tuesday list
arthistory <- tt$artists

# group dataframe by year and nationality
df <- arthistory %>%
  group_by(year, artist_nationality) %>%
  summarise(counts = n())
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

Visualize

Using your processed dataset, create your unique visualization.

```
df %>%
  ggplot() +
  aes(x = year, y = counts, color = artist_nationality) +
  geom_point()
```



Save Image

Save your image for sharing. Be sure to use the `#TidyTuesday` hashtag in your post on twitter!

```
# This will save your most recent plot
ggsave(
  filename = "My TidyTuesday Plot.png",
  device = "png")
```

```
## Saving 6.5 x 4.5 in image
```