Art History Data Anaysis

Tianyi Ma
05.09.2023

In high school I took AP Art History. Although I only got a 3 out of 5 in the exam, which is one of my worst scores, I loved the subject since I get to look at a sea of artworks. Tidy Tuesday has a ton of fun datasets, but this one stood out to me because they actually extracted data from two Art History textbooks, and I think one of them is actually my high school Art History textbook. I'm compelled to analyze it at this point.

Read Dataset

```
library(tidytuesdayR)
## Warning: package 'tidytuesdayR' was built under R version 4.2.3
# tuesdata <- tidytuesdayR::tt_load('2023-01-17')
# tuesdata <- tidytuesdayR::tt_load(2023, week = 03)

# arthistory <- tuesdata$arthistory
```

It seems like the TidyTuesday version of the dataset only has artist.csv. This is a little disappointing since I originally considered the artwork data to be the meat of this dataset. Luckily the github README file for this dataset cited its original publish site, so I went on there and there it is, the instruction of downloading the full dataset. Let's do it.

```
# install.packages("arthistory")
```

Now let's take a quick look of the inside of the `worksgardner` dataset.

```
library("arthistory")
## Warning: package 'arthistory' was built under R version 4.2.3
head(worksgardner, 5)
## # A tibble: 5 × 24
##   artist_name edition_number title_of_work publication_year
page_number_of_image
##   <chr>                <dbl> <chr>                    <dbl> <chr>
## 1 Aaron Doug…              9 Noah's Ark                1991 965
## 2 Aaron Doug…             10 Noah's Ark                1996 1053
## 3 Aaron Doug…             11 Noah's Ark                2001 1030
## 4 Aaron Doug…             12 Noah's Ark                2005 990
## 5 Aaron Doug…             13 Noah's Ark                2009 937
## # ℹ 19 more variables: artist_unique_id <dbl>, artist_nationality <chr>,
## #   artist_gender <chr>, artist_race <chr>, artist_ethnicity <chr>,
## #   height_of_work_in_book <dbl>, width_of_work_in_book <dbl>,
## #   height_of_text <dbl>, width_of_text <dbl>, extra_text_height <dbl>,
## #   extra_text_width <dbl>, area_of_work_in_book <dbl>, area_of_text <dbl>,
## #   extra_text_area <dbl>, total_area_text <dbl>, total_space <dbl>,
## #   page_area <dbl>, space_ratio_per_page <dbl>, book <chr>
```

Getting basic info on the data:

```
str(worksgardner)
## tibble [2,325 × 24] (S3: tbl_df/tbl/data.frame)
```

```
##  $ artist_name           : chr [1:2325] "Aaron Douglas" "Aaron Douglas"
"Aaron Douglas" "Aaron Douglas" ...
##  $ edition_number        : num [1:2325] 9 10 11 12 13 14 15 16 14 15 ...
##  $ title_of_work         : chr [1:2325] "Noah's Ark" "Noah's Ark" "Noah's
Ark" "Noah's Ark" ...
##  $ publication_year      : num [1:2325] 1991 1996 2001 2005 2009 ...
##  $ page_number_of_image  : chr [1:2325] "965" "1053" "1030" "990" ...
##  $ artist_unique_id      : num [1:2325] 1 1 1 1 1 1 1 1 2 2 ...
##  $ artist_nationality    : chr [1:2325] "American" "American" "American"
"American" ...
##  $ artist_gender         : chr [1:2325] "Male" "Male" "Male" "Male" ...
##  $ artist_race           : chr [1:2325] "Black or African American" "Black
or African American" "Black or African American" "Black or African
American" ...
##  $ artist_ethnicity      : chr [1:2325] "Not Hispanic or Latinx" "Not
Hispanic or Latinx" "Not Hispanic or Latinx" "Not Hispanic or Latinx" ...
##  $ height_of_work_in_book: num [1:2325] 11.3 12.1 12.3 12.3 12.8 12.8 12.7
7.9 14 12.8 ...
##  $ width_of_work_in_book : num [1:2325] 8.5 8.9 8.8 8.8 9.3 9.3 9.2 19 10.2
9.2 ...
##  $ height_of_text        : num [1:2325] 14.5 12.4 10.8 15.7 15 18.8 21.2
14.7 4.5 16.2 ...
##  $ width_of_text         : num [1:2325] 8.4 9 9 8.9 9.3 9.3 9.2 13.9 9.3
9.2 ...
##  $ extra_text_height     : num [1:2325] 0 0 0 0 0 0 0 0 9.2 0 ...
##  $ extra_text_width      : num [1:2325] 0 0 0 0 0 0 0 0 8.8 0 ...
##  $ area_of_work_in_book  : num [1:2325] 96 108 108 108 119 ...
##  $ area_of_text          : num [1:2325] 121.8 111.6 97.2 139.7 139.5 ...
##  $ extra_text_area       : num [1:2325] 0 0 0 0 0 ...
##  $ total_area_text       : num [1:2325] 121.8 111.6 97.2 139.7 139.5 ...
##  $ total_space           : num [1:2325] 218 219 205 248 259 ...
##  $ page_area             : num [1:2325] 616 586 677 658 649 ...
##  $ space_ratio_per_page  : num [1:2325] 0.353 0.374 0.303 0.377 0.398 ...
##  $ book                  : chr [1:2325] "gardner" "gardner" "gardner"
"gardner" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ArtistName = col_character(),
##   ..   EditionNumber = col_double(),
##   ..   TitleofWork = col_character(),
##   ..   Year = col_double(),
##   ..   PageNumberofImage = col_character(),
##   ..   ArtistUniqueID = col_double(),
##   ..   ArtistNationality = col_character(),
##   ..   ArtistGender = col_character(),
##   ..   ArtistRace = col_character(),
##   ..   ArtistEthnicity = col_character(),
##   ..   `HeightofWorkinGardner(cm)*` = col_double(),
##   ..   `WidthofWorkinGardner(cm)` = col_double(),
##   ..   `LengthofText(cm)` = col_double(),
##   ..   `WidthofText(cm)` = col_double(),
##   ..   `ExtraTextLength(cm)` = col_double(),
##   ..   `ExtraTextWidth(cm)` = col_double(),
##   ..   `AreaofWorkinGardner(cm^2)` = col_double(),
```

```
##   ..   `AreaofText(cm^2)` = col_double(),
##   ..   `ExtraTextArea(cm^2)` = col_double(),
##   ..   `TotalAreaText(cm^2)` = col_double(),
##   ..   `TotalSpace(cm^2)` = col_double(),
##   ..   `PageArea(cm^2)` = col_double(),
##   ..   SpaceRatioPerPage = col_double(),
##   ..   Book = col_character(),
##   ..   ...25 = col_skip(),
##   ..   ...26 = col_skip(),
##   ..   ...27 = col_skip(),
##   ..   ...28 = col_skip(),
##   ..   ...29 = col_skip(),
##   ..   ...30 = col_skip(),
##   ..   ...31 = col_skip(),
##   ..   ...32 = col_skip(),
##   ..   ...33 = col_skip(),
##   ..   ...34 = col_skip(),
##   ..   ...35 = col_skip(),
##   ..   ...36 = col_skip(),
##   ..   ...37 = col_skip(),
##   ..   ...38 = col_skip(),
##   ..   ...39 = col_skip(),
##   ..   ...40 = col_skip(),
##   ..   ...41 = col_skip(),
##   ..   ...42 = col_skip(),
##   ..   ...43 = col_skip(),
##   ..   ...44 = col_skip(),
##   ..   ...45 = col_skip(),
##   ..   ...46 = col_skip(),
##   ..   ...47 = col_skip(),
##   ..   ...48 = col_skip(),
##   ..   ...49 = col_skip(),
##   ..   ...50 = col_skip(),
##   ..   ...51 = col_skip(),
##   ..   ...52 = col_skip()
##   .. )
```

That is a little disappointing – I was expecting to see descriptive texts associated with each artwork. But this is okay. The size of the artwork and its texts also could be interesting.

Right now, though, we observe that the first 5 rows are all work called “Noah’s Ark” by Aaron Douglas, just in different editions of the book. Because I don’t care about different editions and sizes right now, I’ll simply remove every edition other than the latest one for every author and their artwork. Right now, I want to analyze artist and their artwork info. Later, we can try answer questions like “how many artworks are included in each edition” or “How many artworks are added for each edition”.

The following snippet group row values by column `artist_name` and `title_of_work`, then only retain the last row. Because the edition row values are in descending order, this should do the trick. Checking the first two rows shows that the code is working.

```
library(dplyr)
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
worksgardner_lean <- worksgardner %>%
  group_by(artist_name, title_of_work) %>%
  filter(row_number() == n()) %>%
  ungroup()
head(worksgardner_lean, 2)
## # A tibble: 2 × 24
##   artist_name edition_number title_of_work publication_year
page_number_of_image
##   <chr>                <dbl> <chr>                     <dbl> <chr>
## 1 Aaron Doug…             15 Noah's Ark                 2016 912
## 2 Aaron Doug…             16 From Slavery…              2020 932
## # ℹ 19 more variables: artist_unique_id <dbl>, artist_nationality <chr>,
## #   artist_gender <chr>, artist_race <chr>, artist_ethnicity <chr>,
## #   height_of_work_in_book <dbl>, width_of_work_in_book <dbl>,
## #   height_of_text <dbl>, width_of_text <dbl>, extra_text_height <dbl>,
## #   extra_text_width <dbl>, area_of_work_in_book <dbl>, area_of_text <dbl>,
## #   extra_text_area <dbl>, total_area_text <dbl>, total_space <dbl>,
## #   page_area <dbl>, space_ratio_per_page <dbl>, book <chr>
```

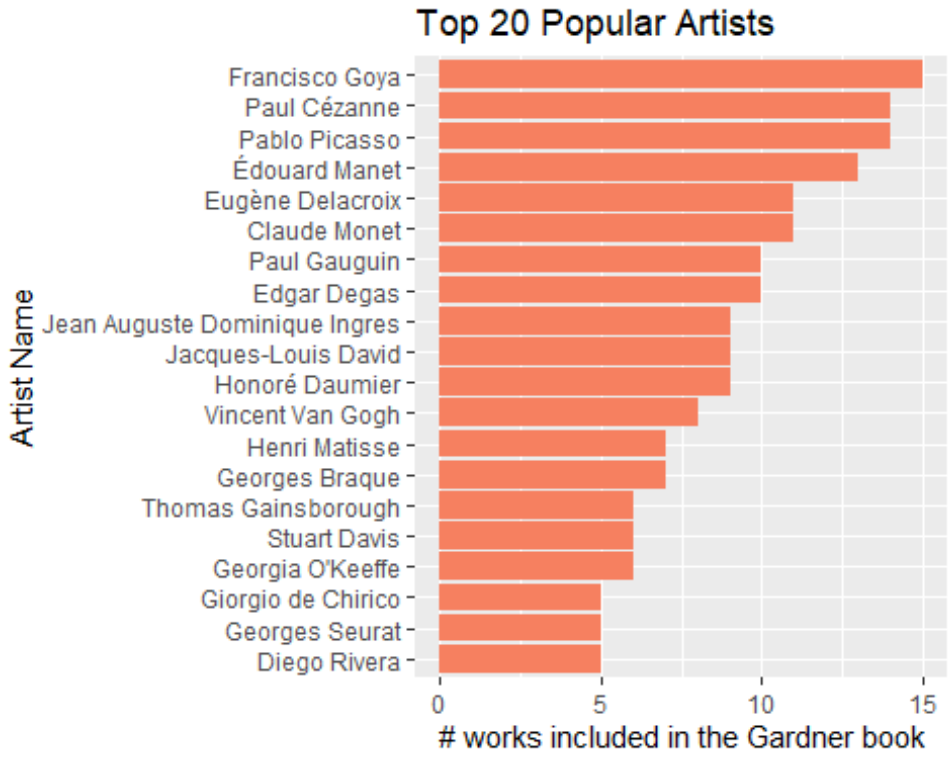
Further looking into some basic statistic in the new dataframe shows that the number of data points shrunk from 2300+ to ~750 artist-artwork pairs.

```
str(worksgardner_lean)
## tibble [746 × 24] (S3: tbl_df/tbl/data.frame)
##  $ artist_name           : chr [1:746] "Aaron Douglas" "Aaron Douglas"
"Adélaïde Labille-Guiard" "Adolphe William Bouguereau" ...
##  $ edition_number        : num [1:746] 15 16 16 6 13 14 16 3 4 10 ...
##  $ title_of_work         : chr [1:746] "Noah's Ark" "From Slavery through
Reconstruction, from Aspects of Negro Life" "Self-Portrait with Two Pupils"
"The Birth of Venus" ...
##  $ publication_year      : num [1:746] 2016 2020 2020 1975 2009 ...
##  $ page_number_of_image  : chr [1:746] "912" "932" "784" "683" ...
##  $ artist_unique_id      : num [1:746] 1 1 2 3 3 4 4 5 5 5 ...
##  $ artist_nationality    : chr [1:746] "American" "American" "French"
"French" ...
##  $ artist_gender         : chr [1:746] "Male" "Male" "Female" "Male" ...
##  $ artist_race           : chr [1:746] "Black or African American" "Black
or African American" "White" "White" ...
##  $ artist_ethnicity      : chr [1:746] "Not Hispanic or Latinx" "Not
Hispanic or Latinx" "Not Hispanic or Latinx" "Not Hispanic or Latinx" ...
##  $ height_of_work_in_book: num [1:746] 12.7 7.9 12.8 10 13.6 8.6 11.2 10.3
8 10.7 ...
##  $ width_of_work_in_book : num [1:746] 9.2 19 9.2 7.3 9.3 14.8 19 12.3 8.5
13.5 ...
##  $ height_of_text        : num [1:746] 21.2 14.7 24.8 9.5 6.9 12.4 5.6 7.7
8 14 ...
##  $ width_of_text         : num [1:746] 9.2 13.9 9.3 7 9.3 9.3 9.3 5.9 6.1 9
```

```
...
##  $ extra_text_height     : num [1:746] 0 0 0 0 0 0 0 0 0 0 ...
##  $ extra_text_width      : num [1:746] 0 0 0 0 0 0 0 0 0 0 ...
##  $ area_of_work_in_book  : num [1:746] 117 150 118 73 126 ...
##  $ area_of_text          : num [1:746] 195 204.3 230.6 66.5 64.2 ...
##  $ extra_text_area       : num [1:746] 0 0 0 0 0 0 0 0 0 0 ...
##  $ total_area_text       : num [1:746] 195 204.3 230.6 66.5 64.2 ...
##  $ total_space           : num [1:746] 312 354 348 140 191 ...
##  $ page_area             : num [1:746] 647 652 652 468 649 ...
##  $ space_ratio_per_page  : num [1:746] 0.482 0.544 0.534 0.298 0.294 ...
##  $ book                  : chr [1:746] "gardner" "gardner" "gardner"
"gardner" ...
```

Now we can answer some question like “who’s the most popular artist according to this textbook?”. We will show this as a bar graph.

```
library('ggplot2')
## Warning: package 'ggplot2' was built under R version 4.2.3
library('forcats')
## Warning: package 'forcats' was built under R version 4.2.3
artist_work_num_rank <- worksgardner_lean %>%
                          count(artist_name, sort = T)

artist_work_num_rank %>%
    mutate(artist_name = fct_reorder(artist_name, n)) %>%
    slice(1:20) %>%
    ggplot(., aes(x=artist_name, y=n)) +
    geom_bar(stat = "identity", fill="#f68060") +
    coord_flip() +
    ggtitle('Top 20 Popular Artists') +
    xlab('Artist Name') +
    ylab('# works included in the Gardner book')
```



This data package also has a corresponding artist dataset, so I thought it would be interesting to look at info related to artists that are featured in this book. Let's load, take a look, and join the table and do some wrangling.

```
head(artists, 3)
## # A tibble: 3 × 14
##   artist_name   edition_number  year artist_nationality
artist_nationality_other
##   <chr>                  <dbl> <dbl> <chr>              <chr>
## 1 Aaron Douglas              9  1991 American           American
## 2 Aaron Douglas             10  1996 American           American
## 3 Aaron Douglas             11  2001 American           American
## # ℹ 9 more variables: artist_gender <chr>, artist_race <chr>,
## #   artist_ethnicity <chr>, book <chr>, space_ratio_per_page_total <dbl>,
## #   artist_unique_id <dbl>, moma_count_to_year <dbl>,
## #   whitney_count_to_year <dbl>, artist_race_nwi <chr>
```

It turns out the same pattern exists for this dataset – multiple entries exist for the same artist because of different editions. So here we first apply the same data transformation to only get the last row, and then join the two tables.

```
artists_lean <- artists %>%
  group_by(artist_name) %>%
  filter(row_number() == n()) %>%
  ungroup()

by <- join_by(artist_name)
```

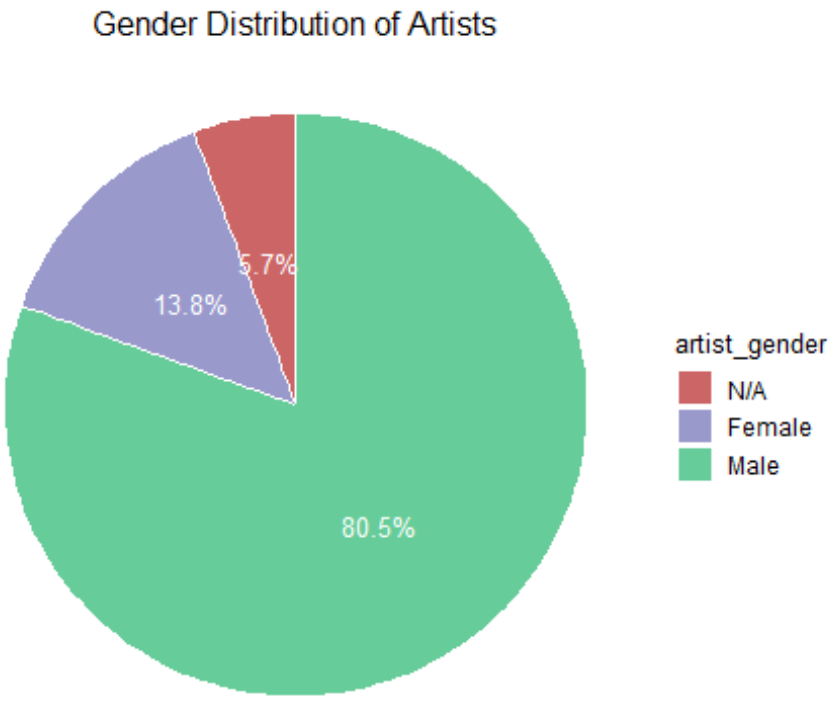
```
artists_lean <- left_join(artist_work_num_rank, artists_lean) %>%
                  rename("num_artworks" = "n")
## Joining with `by = join_by(artist_name)`
head(artists_lean, 3)
## # A tibble: 3 × 15
##   artist_name    num_artworks edition_number  year artist_nationality
##   <chr>                 <int>          <dbl> <dbl> <chr>
## 1 Francisco Goya           15              8  2011 Spanish
## 2 Pablo Picasso            14              8  2011 Spanish
## 3 Paul Cézanne             14              8  2011 French
## # i 10 more variables: artist_nationality_other <chr>, artist_gender <chr>,
## #   artist_race <chr>, artist_ethnicity <chr>, book <chr>,
## #   space_ratio_per_page_total <dbl>, artist_unique_id <dbl>,
## #   moma_count_to_year <dbl>, whitney_count_to_year <dbl>,
## #   artist_race_nwi <chr>
```

Now – the data has attributes like gender, nationality, race, and ethnicity, so obviously we should plot a pie chart for the distribution of each of those attributes.

```
gender_dist <- artists_lean %>%
                group_by(artist_gender) %>%
                summarise(gender_sum=n())
artists_lean %>% ungroup()
## # A tibble: 334 × 15
##    artist_name         num_artworks edition_number  year artist_nationality
##    <chr>                      <int>          <dbl> <dbl> <chr>
##  1 Francisco Goya                15              8  2011 Spanish
##  2 Pablo Picasso                 14              8  2011 Spanish
##  3 Paul Cézanne                  14              8  2011 French
##  4 Édouard Manet                 13              8  2011 French
##  5 Claude Monet                  11              8  2011 French
##  6 Eugène Delacroix              11              8  2011 French
##  7 Edgar Degas                   10              8  2011 French
##  8 Paul Gauguin                  10              8  2011 French
##  9 Honoré Daumier                 9              8  2011 French
## 10 Jacques-Louis David            9              8  2011 French
## # i 324 more rows
## # i 10 more variables: artist_nationality_other <chr>, artist_gender <chr>,
## #   artist_race <chr>, artist_ethnicity <chr>, book <chr>,
## #   space_ratio_per_page_total <dbl>, artist_unique_id <dbl>,
## #   moma_count_to_year <dbl>, whitney_count_to_year <dbl>,
## #   artist_race_nwi <chr>
nationality_dist <- artists_lean %>%
                group_by(artist_nationality) %>%
                summarise(nationality_sum=n())
artists_lean %>% ungroup()
## # A tibble: 334 × 15
##    artist_name         num_artworks edition_number  year artist_nationality
##    <chr>                      <int>          <dbl> <dbl> <chr>
##  1 Francisco Goya                15              8  2011 Spanish
##  2 Pablo Picasso                 14              8  2011 Spanish
##  3 Paul Cézanne                  14              8  2011 French
##  4 Édouard Manet                 13              8  2011 French
##  5 Claude Monet                  11              8  2011 French
##  6 Eugène Delacroix              11              8  2011 French
```

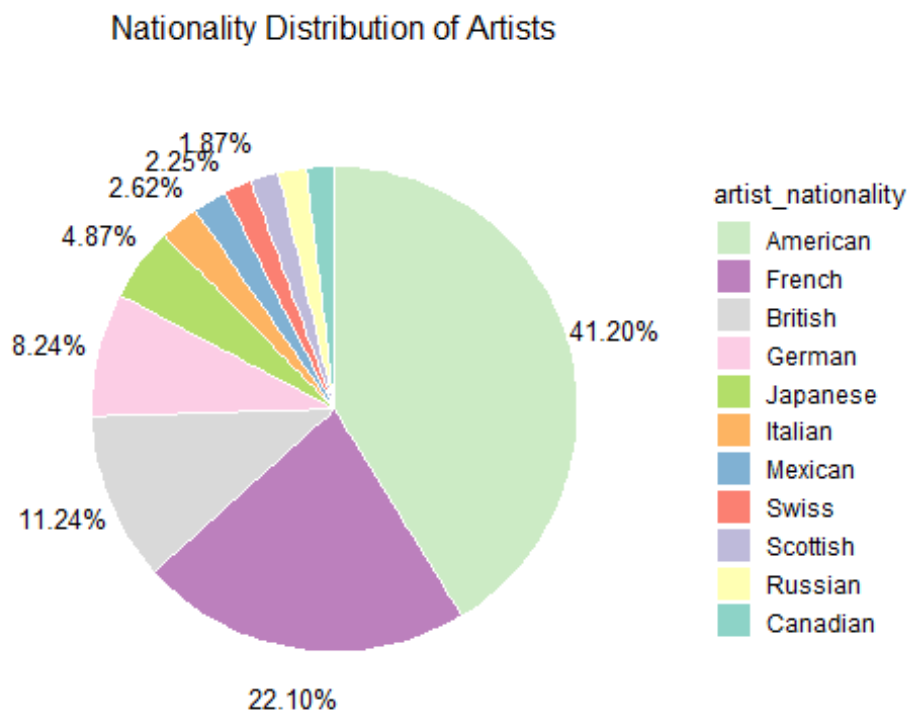
```
##  7 Edgar Degas                   10              8  2011 French
##  8 Paul Gauguin                  10              8  2011 French
##  9 Honoré Daumier                 9              8  2011 French
## 10 Jacques-Louis David            9              8  2011 French
## # i 324 more rows
## # i 10 more variables: artist_nationality_other <chr>, artist_gender <chr>,
## #   artist_race <chr>, artist_ethnicity <chr>, book <chr>,
## #   space_ratio_per_page_total <dbl>, artist_unique_id <dbl>,
## #   moma_count_to_year <dbl>, whitney_count_to_year <dbl>,
## #   artist_race_nwi <chr>
race_dist <- artists_lean %>%
               group_by(artist_race) %>%
               summarise(race_sum=n())
artists_lean %>% ungroup()
## # A tibble: 334 × 15
##    artist_name         num_artworks edition_number  year artist_nationality
##    <chr>                      <int>          <dbl> <dbl> <chr>
##  1 Francisco Goya                15              8  2011 Spanish
##  2 Pablo Picasso                 14              8  2011 Spanish
##  3 Paul Cézanne                  14              8  2011 French
##  4 Édouard Manet                 13              8  2011 French
##  5 Claude Monet                  11              8  2011 French
##  6 Eugène Delacroix              11              8  2011 French
##  7 Edgar Degas                   10              8  2011 French
##  8 Paul Gauguin                  10              8  2011 French
##  9 Honoré Daumier                 9              8  2011 French
## 10 Jacques-Louis David            9              8  2011 French
## # i 324 more rows
## # i 10 more variables: artist_nationality_other <chr>, artist_gender <chr>,
## #   artist_race <chr>, artist_ethnicity <chr>, book <chr>,
## #   space_ratio_per_page_total <dbl>, artist_unique_id <dbl>,
## #   moma_count_to_year <dbl>, whitney_count_to_year <dbl>,
## #   artist_race_nwi <chr>
ethnicity_dist <- artists_lean %>%
               group_by(artist_ethnicity) %>%
               summarise(ethnicity_sum=n())
artists_lean %>% ungroup()
## # A tibble: 334 × 15
##    artist_name         num_artworks edition_number  year artist_nationality
##    <chr>                      <int>          <dbl> <dbl> <chr>
##  1 Francisco Goya                15              8  2011 Spanish
##  2 Pablo Picasso                 14              8  2011 Spanish
##  3 Paul Cézanne                  14              8  2011 French
##  4 Édouard Manet                 13              8  2011 French
##  5 Claude Monet                  11              8  2011 French
##  6 Eugène Delacroix              11              8  2011 French
##  7 Edgar Degas                   10              8  2011 French
##  8 Paul Gauguin                  10              8  2011 French
##  9 Honoré Daumier                 9              8  2011 French
## 10 Jacques-Louis David            9              8  2011 French
## # i 324 more rows
## # i 10 more variables: artist_nationality_other <chr>, artist_gender <chr>,
## #   artist_race <chr>, artist_ethnicity <chr>, book <chr>,
## #   space_ratio_per_page_total <dbl>, artist_unique_id <dbl>,
```

```
## #   moma_count_to_year <dbl>, whitney_count_to_year <dbl>,
## #   artist_race_nwi <chr>
library("scales")
## Warning: package 'scales' was built under R version 4.2.3
gender_dist <- gender_dist %>%
                  arrange(desc(gender_sum)) %>%
                  mutate(prop = gender_sum/sum(gender_sum) *100) %>%
                  mutate(ypos = cumsum(prop)- 0.5*prop) %>%
                  ungroup()

gender_dist_graph <- gender_dist %>%
                      mutate(artist_gender = fct_reorder(artist_gender, prop))
%>%
                      ggplot(., aes(x="", y=prop, fill=artist_gender)) +
                        geom_bar(stat="identity", width=1, color="white") +
                        coord_polar("y", start=0) +
                        scale_fill_manual(values=c("#CC6666", "#9999CC",
"#66CC99")) +
                        theme_void() +
                        geom_text(aes(y = ypos, label = percent(prop/100)),
color = "white", size=3.5) +
                        ggtitle('Gender Distribution of Artists') +
                        theme(plot.title = element_text(hjust = 0.5))

gender_dist_graph <- gender_dist_graph +
                        theme(legend.key.size = unit(0.5, 'cm'),
                              legend.title = element_text(size=10),
                              legend.text = element_text(size=9),
                              plot.title=element_text(size=12))

gender_dist_graph
```



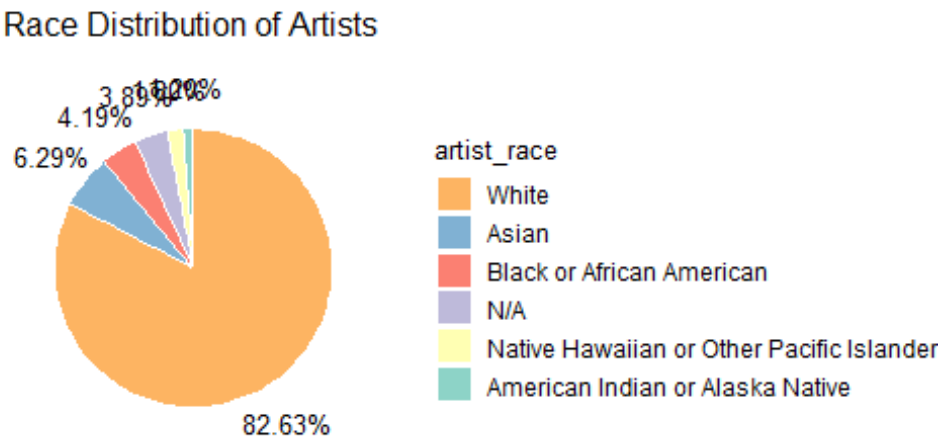

```
library("tidyverse")
## Warning: package 'tidyverse' was built under R version 4.2.3
## Warning: package 'tibble' was built under R version 4.2.3
## Warning: package 'tidyr' was built under R version 4.2.3
## Warning: package 'readr' was built under R version 4.2.3
## Warning: package 'lubridate' was built under R version 4.2.3
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ lubridate 1.9.2     ✔ stringr   1.5.0
## ✔ purrr     1.0.1     ✔ tibble    3.2.1
## ✔ readr     2.1.4     ✔ tidyr     1.3.0
## ── Conflicts ────────────────────────────────────────── 
tidyverse_conflicts() ──
## ✖ readr::col_factor() masks scales::col_factor()
## ✖ purrr::discard()    masks scales::discard()
## ✖ dplyr::filter()     masks stats::filter()
## ✖ dplyr::lag()        masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force
all conflicts to become errors
library("ggrepel")
## Warning: package 'ggrepel' was built under R version 4.2.3
nationality_dist <- nationality_dist[nationality_dist$artist_nationality !=
"N/A",]

nationality_dist <- nationality_dist %>%
                      mutate(artist_nationality = fct_lump(artist_nationality,
n = 11, w = nationality_sum, other_level = "Others")) %>%
                      group_by(artist_nationality) %>%
```

```
                      summarize(nationality_sum = sum(nationality_sum)) %>%
                      arrange(desc(nationality_sum))

nationality_dist <- subset(nationality_dist, artist_nationality != "Others")

nationality_dist <- nationality_dist %>%
                      mutate(prop = nationality_sum/sum(nationality_sum) *100)
%>%
                      mutate(ypos = cumsum(prop) - 0.5*prop)

labels <- c(percent(nationality_dist$prop/100)[1:8], c("", "", ""))

nationality_dist_graph <- nationality_dist %>%
                      mutate(artist_nationality =
fct_reorder(artist_nationality, prop)) %>%
                      ggplot(., aes(x="", y=prop, fill=artist_nationality)) +
                        geom_bar(stat="identity", width=1, color="white") +
                        coord_polar("y", start=0) +
                        scale_fill_brewer(palette = "Set3") +
                        theme_void() +
                        geom_text(aes(y = ypos, label = labels), size=3.5,
show.legend = F, nudge_x = 0.7) +
                        ggtitle('Nationality Distribution of Artists') +
                        theme(plot.title = element_text(hjust = 0.5)) +
                        guides(fill = guide_legend(reverse=T))

nationality_dist_graph <- nationality_dist_graph +
                        theme(legend.key.size = unit(0.5, 'cm'),
                              legend.title = element_text(size=10),
                              legend.text = element_text(size=9),
                              plot.title=element_text(size=12))
nationality_dist_graph
```



Now for the race data:

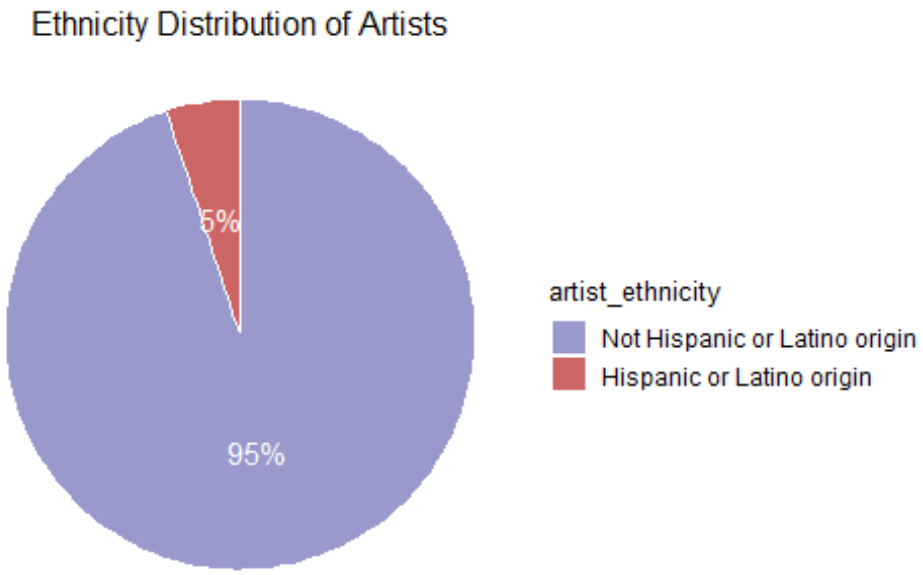
```
race_dist <- race_dist %>%
              arrange(desc(race_sum)) %>%
              mutate(prop = race_sum/sum(race_sum) *100) %>%
              mutate(ypos = cumsum(prop) - 0.5*prop)

race_dist_graph <- race_dist %>%
                    mutate(artist_race = fct_reorder(artist_race, prop)) %>%
                    ggplot(., aes(x="", y=prop, fill=artist_race)) +
                      geom_bar(stat="identity", width=1, color="white") +
                      coord_polar("y", start=0) +
                      scale_fill_brewer(palette = "Set3") +
                      theme_void() +
                      geom_text(aes(y = ypos, label = percent(prop/100)),
size=3.5, show.legend = F, nudge_x = 0.8) +
                      ggtitle('Race Distribution of Artists') +
                      theme(plot.title = element_text(hjust = 0.5)) +
                      guides(fill = guide_legend(reverse=T))

race_dist_graph <- race_dist_graph +
                        theme(legend.key.size = unit(0.5, 'cm'),
                              legend.title = element_text(size=10),
                              legend.text = element_text(size=9),
                              plot.title=element_text(size=12))
race_dist_graph
```



Ethnicity distribution:

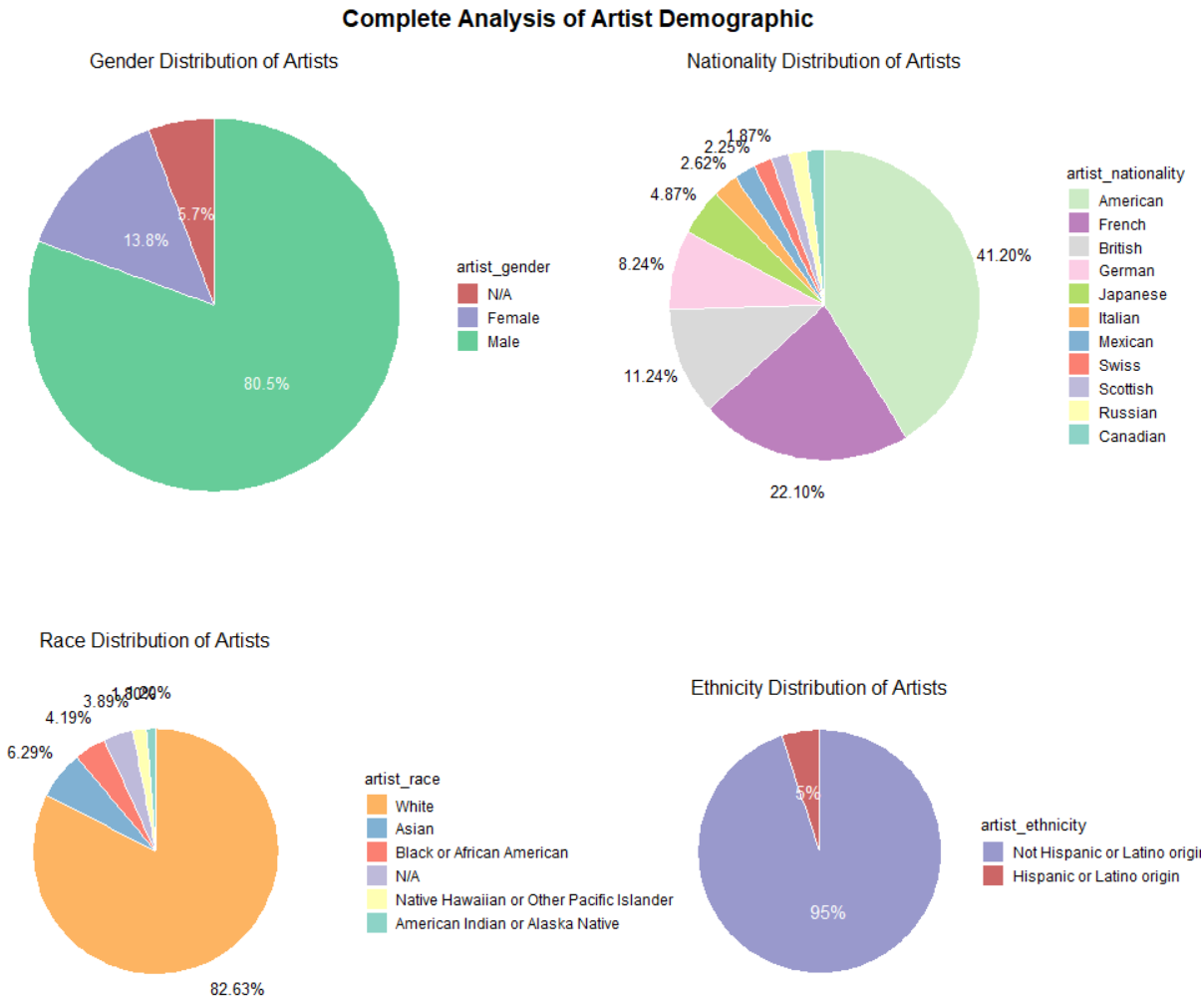
```
ethnicity_dist <- ethnicity_dist[ethnicity_dist$artist_ethnicity != "NA",]
ethnicity_dist <- na.omit(ethnicity_dist)

ethnicity_dist <- ethnicity_dist %>%
                    arrange(desc(ethnicity_sum)) %>%
                    mutate(prop = ethnicity_sum/sum(ethnicity_sum) *100) %>%
                    mutate(ypos = cumsum(prop) - 0.5*prop)

ethnicity_dist_graph <- ethnicity_dist %>%
                          mutate(artist_ethnicity =
fct_reorder(artist_ethnicity, prop)) %>%
                          ggplot(., aes(x="", y=prop,
fill=artist_ethnicity)) +
                          geom_bar(stat="identity", width=1, color="white")
+
                          coord_polar("y", start=0) +
                          scale_fill_manual(values=c("#CC6666", "#9999CC"))
+
                          theme_void() +
                          geom_text(aes(y = ypos, label =
percent(prop/100)), size=4, color = "white") +
                          ggtitle('Ethnicity Distribution of Artists') +
                          theme(plot.title = element_text(hjust = 0.5)) +
                          guides(fill = guide_legend(reverse=T))

ethnicity_dist_graph <- ethnicity_dist_graph +
```

```
                        theme(legend.key.size = unit(0.5, 'cm'),
                              legend.title = element_text(size=10),
                              legend.text = element_text(size=9),
                              plot.title=element_text(size=12))

ethnicity_dist_graph
```



Putting this together:

```
library(cowplot)
## Warning: package 'cowplot' was built under R version 4.2.3
##
## Attaching package: 'cowplot'
## The following object is masked from 'package:lubridate':
##
##     stamp
title <- ggdraw() + draw_label("Complete Analysis of Artist Demographic",
fontface='bold')
top_plot <- plot_grid(gender_dist_graph, nationality_dist_graph, ncol = 2,
labels = c("", ""), rel_heights = c(0.5, 2))
bottom_plot <- plot_grid(race_dist_graph, ethnicity_dist_graph, ncol = 2,
labels = c("", ""), rel_widths = c(1, 0.75))

plot_grid(title, top_plot, bottom_plot, nrow = 3, labels = c("", "", ""),
          rel_heights = c(0.1, 1, 1.2))
```


Complete Analysis of Artist Demographic
Gender Distribution of Artists
5.7%
13.8%
80.5%
artist_gender
N/A
Female
Male
Nationality Distribution of Artists
1.87%
2.25%
2.62%
4.87%
8.24%
41.20%
11.24%
22.10%
artist_nationality
American
French
British
German
Japanese
Italian
Mexican
Swiss
Scottish
Russian
Canadian
Race Distribution of Artists
3.89%
4.19%
6.29%
82.63%
artist_race
White
Asian
Black or African American
N/A
Native Hawaiian or Other Pacific Islander
American Indian or Alaska Native
Ethnicity Distribution of Artists
5%
95%
artist_ethnicity
Not Hispanic or Latino origi
Hispanic or Latino origin