

So while the curating of new data is dead, it is still never the less cool to explore this data and learn about the Tate Museum.

Some domain knowledge- What is the Tate Museum?



Source: <https://www.tate.org.uk/whats-on/tate-st-ives/collection-route/modern-art-and-st-ives-and-haegue-yang>

While for some this is may be obvious (shout out to my British readers!), a Canadian like myself never heard of this museum. So before we run into exploring this data set, lets look [Tate's website](#) and see what it can tell us:

"When Tate first opened its doors to the public in 1897 it had just one site, displaying a small collection of British artworks. Today we have four major sites and the national collection of British art from 1500 to the present day and international modern and contemporary art, which includes nearly 70,000 artworks."

The [readme.md](#) file from the Tidy Tuesday repository tells us a little more about the data.

"Here we present the metadata for around 70,000 artworks that Tate owns or jointly owns with the National Galleries of Scotland as part of ARTIST ROOMS. Metadata for around 3,500 associated artists is also included."

With this context, we're ready to start our analysis.

Our Questions

What good is an analysis without having some questions to ask? Lets seek to answer some of the following questions.

- Tate's website says that [they have art dating back from the 1500's](#). What is the distribution of the age of the artwork as of October 2014?
- What were the popular mediums of artistic expression since the 16th century?
- What is the gender ratio of the featured artists at Tate?

While these questions are pretty basic, I think there is a lot to discover from these insights about the Tate museum and art history in general.

Lets get into it!

Our data

Lets get started with the standard procedure for loading and getting a general overview of our data set.

```
# Run once
library(tidyverse)
library(ggthemes) #Gotta make some pretty pictures.

tuesdata<-tidytuesdayR::tt_load('2021-01-12')

##
## Downloading file 1 of 2: `artists.csv`
## Downloading file 2 of 2: `artwork.csv`

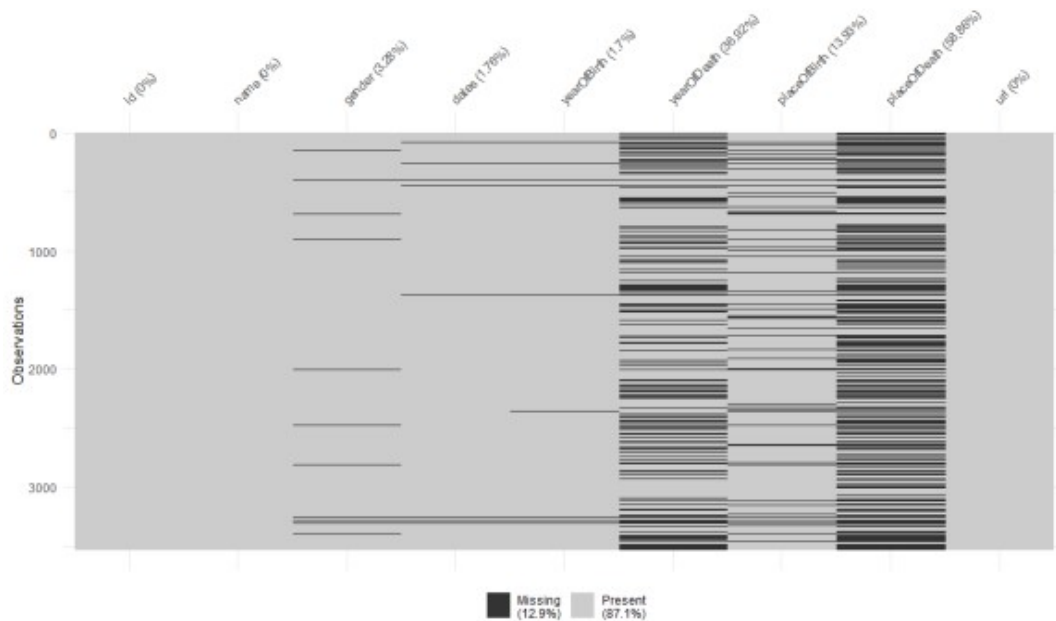
artists<- tuesdata$artists
artwork<- tuesdata$artwork

glimpse(artists)

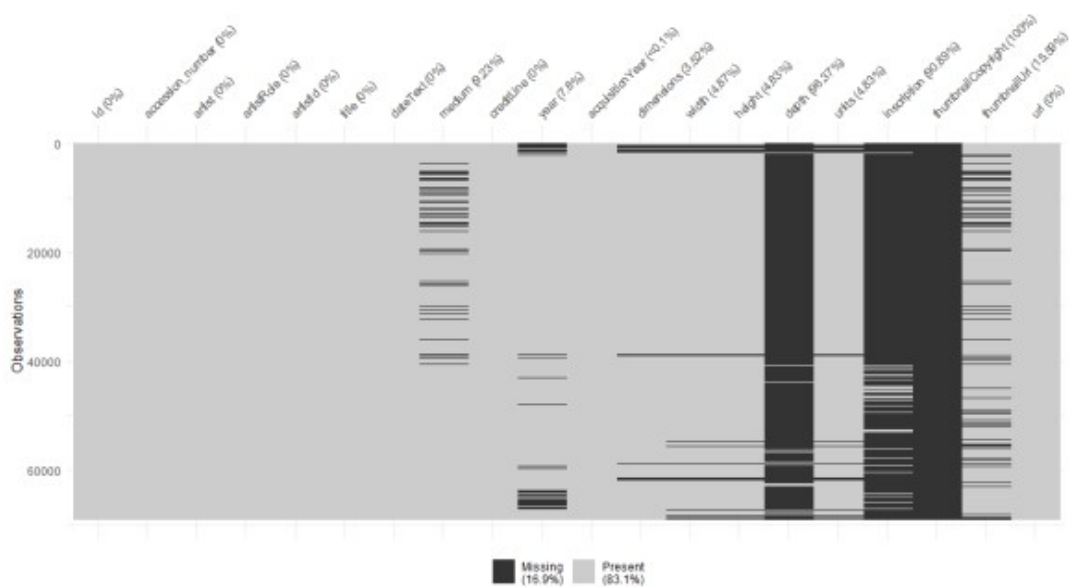
## Rows: 3,532
## Columns: 9
## $ id          10093, 0, 2756, 1, 622, 2606, 9550, 623, 624, 625, 2411,
626, 627, 628, 629, 630, 2608, 631, 632, 633, 2, 3098, 8...
## $ name        "Abakanowicz, Magdalena", "Abbey, Edwin Austin", "Abbott,
Berenice", "Abbott, Lemuel Francis", "Abrahams, Ivor", ...
## $ gender       "Female", "Male", "Female", "Male", "Male", "Male", "Male",
"Female", "Male", "Male", "Male", "Male", "Male", "Male", "Ma...
## $ dates        "born 1930", "1852-1911", "1898-1991", "1760-1803", "born
1935", "1964-1993", "born 1967", "born 1940", "1947-201...
## $ yearOfBirth  1930, 1852, 1898, 1760, 1935, 1964, 1967, 1940, 1947, 1938,
1728, 1868, 1927, 1917, 1878, 1895, 1904, 1927, 1912,...
## $ yearOfDeath  NA, 1911, 1991, 1803, NA, 1993, NA, NA, 2014, NA, 1792,
1947, 2005, 1984, 1966, 1949, 1995, 1987, 1976, 1991, 184...
## $ placeOfBirth "Polska", "Philadelphia, United States", "Springfield,
United States", "Leicestershire, United Kingdom", "Wigan, ...
## $ placeOfDeath NA, "London, United Kingdom", "Monson, United States",
"London, United Kingdom", NA, "Paris, France", NA, NA, NA,...
## $ url          "http://www.tate.org.uk/art/artists/magdalena-abakanowicz-10093",
"http://www.tate.org.uk/art/artists/edwin-austi..."

glimpse(artwork)

## Rows: 69,201
## Columns: 20
## $ id          1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043,
1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052,...
```

```
vis_miss(artwork, warn_large_data = FALSE)
```



There is no missing data present in our data set as far as the `id` and `artistId` fields in the `artwork` and `artists` tables, so a full join with these fields should be fine.

Right?

Wrong.

```
artDataOJ <- artwork %>%
  full_join(artists, by = c("artistId" = "id"))
```

```
glimpse(artDataOJ)
```

```
## Rows: 69,395
## Columns: 28
## $ id      1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043,
1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, ...
```

```
## $ accession_number "A00001", "A00002", "A00003", "A00004", "A00005",  
"A00006", "A00007", "A00008", "A00009", "A00010", "A00011...  
## $ artist "Blake, Robert", "Blake, Robert", "Blake, Robert",  
"Blake, Robert", "Blake, William", "Blake, William", "Bl...  
## $ artistRole "artist", "artist", "artist", "artist", "artist",  
"artist", "artist", "artist", "artist", "artist", "artist...  
## $ artistId 38, 38, 38, 38, 39, 39, 39, 39, 39, 39, 39, 39,  
39, 39, 39, 39, 39, 39, 39, 39, 39, 39, 39, 39,...  
## $ title "A Figure Bowing before a Seated Old Man with his Arm  
Outstretched in Benediction. Verso: Indecipherable Sk...  
## $ dateText "date not known", "date not known", "?c.1785", "date  
not known", "1826-7, reprinted 1892", "1826-7, reprint...  
## $ medium "Watercolour, ink, chalk and graphite on paper.  
Verso: graphite on paper", "Graphite on paper", "Graphite o...  
## $ creditLine "Presented by Mrs John Richmond 1922", "Presented by  
Mrs John Richmond 1922", "Presented by Mrs John Richmo...  
## $ year NA, NA, 1785, NA, 1826, 1826, 1826, 1826, 1826, 1826,  
1826, 1828, 1825, 1825, 1825, 1825, 1825, 1825,...  
## $ acquisitionYear 1922, 1922, 1922, 1922, 1919, 1919, 1919, 1919,  
1919, 1919, 1919, 1919, 1919, 1919, 1919, 1919,...  
## $ dimensions "support: 394 x 419 mm", "support: 311 x 213 mm",  
"support: 343 x 467 mm", "support: 318 x 394 mm", "image:...  
## $ width 394, 311, 343, 318, 243, 240, 242, 246, 241, 243,  
236, 184, 197, 197, 200, 198, 198, 198, 199, 198, 198, 19...  
## $ height 419, 213, 467, 394, 335, 338, 334, 340, 335, 340,  
340, 150, 151, 153, 152, 152, 153, 153, 150, 152, 152, 15...  
## $ depth NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,  
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ units "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm",  
"mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm"...  
## $ inscription NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,  
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ thumbnailCopyright NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,  
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ thumbnailUrl "http://www.tate.org.uk/art/images/work/A/A00/A00001\_8.jpg",  
"http://www.tate.org.uk/art/images/work/A/A00/...  
## $ url.x "http://www.tate.org.uk/art/artworks/blake-a-figure-bowing-before-  
a-seated-old-man-with-his-arm-outstretche...  
## $ name "Blake, Robert", "Blake, Robert", "Blake, Robert",  
"Blake, Robert", "Blake, William", "Blake, William", "Bl...  
## $ gender "Male", "Male", "Male", "Male", "Male", "Male",  
"Male", "Male", "Male", "Male", "Male", "Male", "Male", "Ma...  
## $ dates "1762-1787", "1762-1787", "1762-1787", "1762-1787",  
"1757-1827", "1757-1827", "1757-1827", "1757-1827", "17...  
## $ yearOfBirth 1762, 1762, 1762, 1762, 1757, 1757, 1757, 1757, 1757,  
1757, 1757, 1757, 1757, 1757, 1757, 1757, 1757,...  
## $ yearOfDeath 1787, 1787, 1787, 1787, 1827, 1827, 1827, 1827, 1827,  
1827, 1827, 1827, 1827, 1827, 1827, 1827, 1827,...  
## $ placeOfBirth "London, United Kingdom", "London, United Kingdom",  
"London, United Kingdom", "London, United Kingdom", "Lo...  
## $ placeOfDeath "London, United Kingdom", "London, United Kingdom",  
"London, United Kingdom", "London, United Kingdom", "Lo...  
## $ url.y "http://www.tate.org.uk/art/artists/robert-blake-38",  
"http://www.tate.org.uk/art/artists/robert-blake-38",...
```

Seeing that our row size has increased by 194 rows is a little suspicious. What's going on here?

Lets look at an inner join between these tables.

```
artDataIJ<-artwork %>%
  inner_join(artists,by = c("artistId"="id"))

glimpse(artDataIJ)

## Rows: 69,195
## Columns: 28
## $ id          1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043,
1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052,...
## $ accession_number "A00001", "A00002", "A00003", "A00004", "A00005",
"A00006", "A00007", "A00008", "A00009", "A00010", "A00011...
## $ artist         "Blake, Robert", "Blake, Robert", "Blake, Robert",
"Blake, Robert", "Blake, William", "Blake, William", "Bl...
## $ artistRole      "artist", "artist", "artist", "artist", "artist",
"artist", "artist", "artist", "artist", "artist", "artist...
## $ artistId        38, 38, 38, 38, 39, 39, 39, 39, 39, 39, 39, 39, 39,
39, 39, 39, 39, 39, 39, 39, 39, 39, 39, 39, 39, 39,...
## $ title           "A Figure Bowing before a Seated Old Man with his Arm
Outstretched in Benediction. Verso: Indecipherable Sk...
## $ dateText        "date not known", "date not known", "?c.1785", "date
not known", "1826-7, reprinted 1892", "1826-7, reprint...
## $ medium          "Watercolour, ink, chalk and graphite on paper.
Verso: graphite on paper", "Graphite on paper", "Graphite o...
## $ creditLine       "Presented by Mrs John Richmond 1922", "Presented by
Mrs John Richmond 1922", "Presented by Mrs John Richmo...
## $ year            NA, NA, 1785, NA, 1826, 1826, 1826, 1826, 1826, 1826,
1826, 1828, 1825, 1825, 1825, 1825, 1825, 1825, 1825,...
## $ acquisitionYear  1922, 1922, 1922, 1922, 1919, 1919, 1919, 1919, 1919,
1919, 1919, 1919, 1919, 1919, 1919, 1919, 1919, 1919,...
## $ dimensions      "support: 394 x 419 mm", "support: 311 x 213 mm",
"support: 343 x 467 mm", "support: 318 x 394 mm", "image:...
## $ width           394, 311, 343, 318, 243, 240, 242, 246, 241, 243,
236, 184, 197, 197, 200, 198, 198, 198, 199, 198, 198, 19...
## $ height          419, 213, 467, 394, 335, 338, 334, 340, 335, 340,
340, 150, 151, 153, 152, 152, 153, 153, 150, 152, 152, 15...
## $ depth           NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ units           "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm",
"mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm", "mm",...
## $ inscription     NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ thumbnailCopyright NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ thumbnailUrl     "http://www.tate.org.uk/art/images/work/A/A00/A00001\_8.jpg",
"http://www.tate.org.uk/art/images/work/A/A00/...
## $ url.x            "http://www.tate.org.uk/art/artworks/blake-a-figure-bowing-before-
a-seated-old-man-with-his-arm-outstretche...
## $ name            "Blake, Robert", "Blake, Robert", "Blake, Robert",
"Blake, Robert", "Blake, William", "Blake, William", "Bl...
## $ gender          "Male", "Male", "Male", "Male", "Male", "Male", "Male",
"Male", "Male", "Male", "Male", "Male", "Male", "Ma...
## $ dates           "1762-1787", "1762-1787", "1762-1787", "1762-1787",
"1757-1827", "1757-1827", "1757-1827", "1757-1827", "17...
```

```
## $ yearOfBirth      1762, 1762, 1762, 1762, 1757, 1757, 1757, 1757, 1757,
1757, 1757, 1757, 1757, 1757, 1757, 1757, 1757, 1757, ...
## $ yearOfDeath      1787, 1787, 1787, 1787, 1827, 1827, 1827, 1827,
1827, 1827, 1827, 1827, 1827, 1827, 1827, 1827, 1827, ...
## $ placeOfBirth      "London, United Kingdom", "London, United Kingdom",
"London, United Kingdom", "London, United Kingdom", "Lo...
## $ placeOfDeath      "London, United Kingdom", "London, United Kingdom",
"London, United Kingdom", "London, United Kingdom", "Lo...
## $ url.y             "http://www.tate.org.uk/art/artists/robert-blake-38",
"http://www.tate.org.uk/art/artists/robert-blake-38", ...
```

Well, the inner join gives us 194 rows less than the original. Lets look at the rows not present in the inner join.

```
# went back to doing things in a more base R fashion because `filter()` was
more confusing in this case.
```

```
addedValues<-artDataOJ[!c(artDataOJ$artistId %in% artDataIJ$artistId),]
glimpse(addedValues)
```

```
## Rows: 200
## Columns: 28
## $ id              6652, 16597, 138, 88193, 88194, 88195, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ accession_number "N04252", "T05029", "T07923", "T12066", "T12067",
"T12068", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ artist          "?British School", "Hullmandel, Charles", "Kabakov,
Ilya", "The Leach Pottery (St. Ives, UK)", "The Leach P...
## $ artistRole       "artist", "artist", "artist", "artist", "artist",
"artist", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ artistId        19232, 5265, 3462, 12951, 12951, 12951, 3067, 14647,
5221, 15135, 11604, 12604, 14648, 2714, 13866, 18070, ...
## $ title            "Portrait of a Gentleman, probably of the West
Family", "Portrait of J.M.W. Turner ('The Fallacy of Hope'),...
## $ dateText         "1545-60", "published 1851", "1990", "c.1955",
"c.1955", "c.1955", NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ medium           "Oil paint on oak", "Lithograph on paper", "Wooden
construction, 9 doors, wooden ceiling props, 24 light bu...
## $ creditLine        "Presented in memory of R.S. Holford and Sir George
Holford by nine members of their family 1927", "Present...
## $ year             1545, 1851, 1990, 1955, 1955, 1955, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ acquisitionYear   1927, 1988, 2003, 2005, 2005, 2005, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ dimensions        "support: 1330 x 785 mm frame: 1533 x 988 x 86 mm",
"image: 328 x 225 mm", NA, "object: 100 x 230 x 230 mm"...
## $ width            1330, 328, NA, 100, 70, 70, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ height            785, 225, NA, 230, 145, 145, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ depth            NA, NA, NA, 230, 145, 145, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ units             "mm", "mm", NA, "mm", "mm", "mm", NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ inscription      NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
```

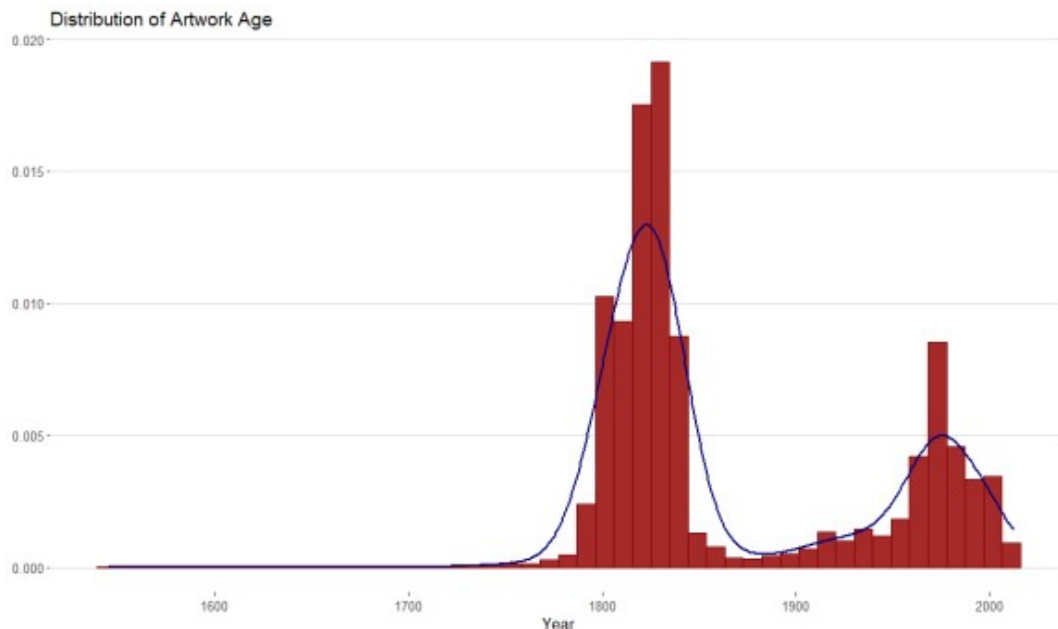
```
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...  
## $ thumbnailCopyright    NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,  
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...  
## $ thumbnailUrl          "http://www.tate.org.uk/art/images/work/N/N04/N04252_8.jpg",  
"http://www.tate.org.uk/art/images/work/T/T05/...  
## $ url.x                 "http://www.tate.org.uk/art/artworks/british-school-portrait-of-  
a-gentleman-probably-of-the-west-family-n04...  
## $ name                  NA, NA, NA, NA, NA, NA, "Albers, Anni", "Amzalag,  
Michael", "Anonymous", "Ant Farm (Chip Lord, born 1944, D...  
## $ gender                NA, NA, NA, NA, NA, NA, "Female", "Male", NA, "Male",  
"Female", "Male", "Male", "Male", "Female", NA, NA, "...  
## $ dates                 NA, NA, NA, NA, NA, NA, "1899-1994", "born 1967", NA,  
NA, "born 1935", "born 1978", "born 1968", "1727-1780...  
## $ yearOfBirth           NA, NA, NA, NA, NA, NA, 1899, 1967, NA, NA, 1935,  
1978, 1968, 1727, 1967, 1740, 1910, 1923, 1959, 1973, 197...  
## $ yearOfDeath           NA, NA, NA, NA, NA, NA, 1994, NA, NA, NA, NA, NA, NA,  
1780, NA, 1799, 1997, 1998, NA, NA, NA, NA, 1828, 180...  
## $ placeOfBirth          NA, NA, NA, NA, NA, NA, "Berlin, Deutschland",  
"Paris, France", NA, NA, "New York, United States", "Buffalo...  
## $ placeOfDeath          NA, NA, NA, NA, NA, NA, "Orange, United States", NA,  
NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ url.y                 NA, NA, NA, NA, NA, NA, "http://www.tate.org.uk/art/artists/anni-  
albers-3067", "http://www.tate.org.uk/art/..."
```

While there is a pretty good job done with the data, clearly, it hasn't been fully audited, or we weren't informed of this issue. Hence the additional 194 rows.

“The dataset in this repository was last updated in October 2014. Tate has no plans to resume updating this repository[...].”

Age Distribution of Art.

```
ggplot(data=artwork,mapping=aes(x=year,y=..density..))+
  theme_hc()+
  geom_histogram(bins = 50,fill="#A52A2A",color="#800000")+
  geom_density(size=0.73,color="#00007F",alpha=0.1,adjust=2,position="stack")+
  labs(x="Year",y=NULL)+
  ggtitle("Distribution of Artwork Age")
```

This definitely looks like more of a bi-modal distribution with large amount of artwork from the first half of the 19th century and the second half of the 20th century.

After a little bit of research, here are some of my proposed reasons for this:

1. The initial art donations to the Museum were from the large collection of Henry Tate; It is likely the large amount of art from the early 19th century was the majority of Henry's personal collection.
2. During the second half of the 20th century, [Contemporary art rose to popularity](#). This together with the social movements of the second half of the 20th century is probably what prompted the spike in purchases and/or submissions of this artwork to the Museum.
3. It's very likely that most of the art here has been donated in some way, but I want to see where most of the purchases lie over time. The museum received significant financial support from the Duveen Family ([who are listed as notable donors of the art gallery](#)) when Tate first opened in 1897. My initial guess is that the Duveen Family was supporting the gallery in the late 19th Century and the early 20th century.
4. The Tate Members ([see the history of Tate webpage](#)) acquired around 400 pieces of art for the museum, so I expect that there will be more purchases also present from 1957 onward.

To verify these claims, lets examine the credit lines of the artwork. After a quick scroll through with `View(artwork)` I see that here are some of the key words that are important to look at.

- **Purchased**
- **Acquired**
- **Presented**
- **Accepted**
- **Bequeathed** and
- **Transferred**

I'm sure there are other terms. Yet, for now lets use these as it *appears* that most of the data uses these terms in the credit line.

Lets get into extracting these terms and classifying them.

(Warning- I'm busting our Base R here because its more intuitive, if you can do this with `mutate()`, comment below so I can check it out!)

```
library(stringr)
bought<- c("Purchased","Acquired")
donated <- c("Presented","Accepted","Bequeathed","Transferred")
```

```
# a simplified field telling us what art has been bought and what has been
donated
boughtIndex<-str_detect(artwork$creditLine,paste(bought,collapse = "|"))
donatedIndex<-str_detect(artwork$creditLine,paste(donated,collapse = "|"))

# I get warnings and it is annoying- but supposedly this is a bug. I'm hiding
it in my markdown output.
artwork<-artwork %>% as.data.frame()
artwork$acquisitionMethod[boughtIndex]<-"Bought"
artwork$acquisitionMethod[donatedIndex]<-"Donated"
```

With that done, lets look at how many NA values we have in comparison with what we have done.

```
sum(is.na(artwork$acquisitionMethod))
```

```
## [1] 12
```

```
sum(is.na(artwork$creditLine))
```

```
## [1] 3
```

Wow! Not bad. With a little bit of exploring by looking at the difference we will add some patterns to categorize 8 of the credit lines for completeness.

One of them we will not categorize because the credit line reads the following.

"Partial purchase and partial gift from the American Fund for the Tate Gallery, courtesy of Julie and Lawrence Salander in memory of John Constable (1776-1837), Daphne Reynolds (1918-2002), Evelyn Joll (1925-2001) and Leslie Parris (1941-2000), 2012. With additional funds provided by Tate Patrons."

This is something which cannot be really classified as being bought or a donation. So for our case we will leave it out.

```
bought<- c("Purchased","Acquired","Exchanged","uncovered")
donated <- c("Presented",
            "Accepted",
            "Bequeathed",
            "Transferred",
            "Gift",
            "Commissioned",
            "Given")
```

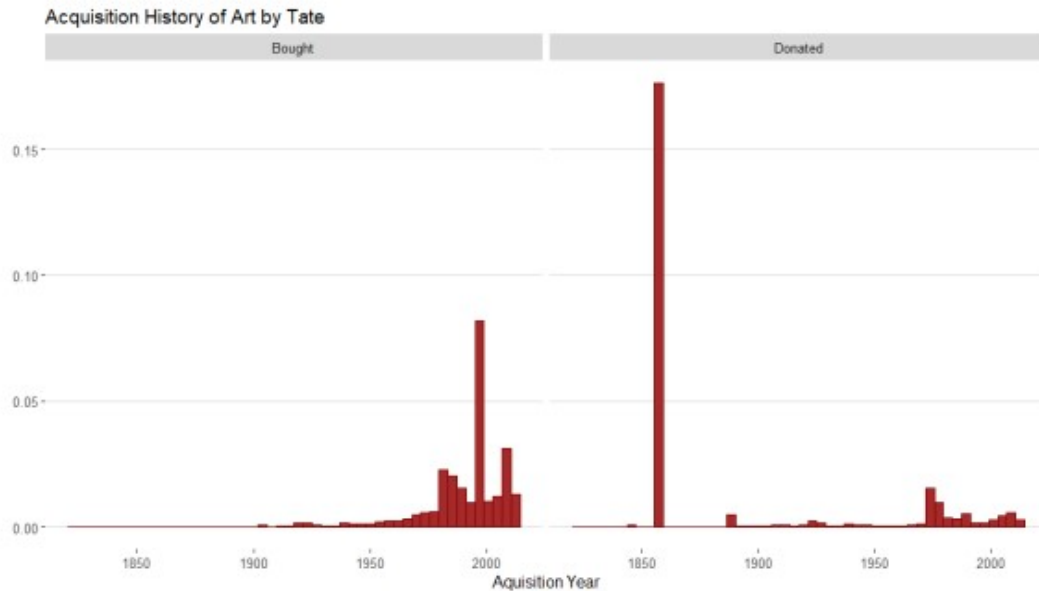
```
# a simplified field telling us what art has been bought and what has been
donated
boughtIndex<-str_detect(artwork$creditLine,paste(bought,collapse = "|"))
donatedIndex<-str_detect(artwork$creditLine,paste(donated,collapse = "|"))

artwork$acquisitionMethod[boughtIndex] <- "Bought"
artwork$acquisitionMethod[donatedIndex] <- "Donated"
```

Now, lets look at the graphs to see if our hypotheses are true. Note that we are now looking at the the

acquisition year of the artwork now.

```
ggplot(data=artwork %>% filter(!is.na(acquisitionMethod)))+  
  theme_hc()+  
  geom_histogram(mapping=aes(x=acquisitionYear, y=..density..),bins = 50,  
fill="#A52A2A", color="#800000")+  
  facet_wrap(~acquisitionMethod)+  
  labs(x="Aquisition Year",y="")+  
  ggtitle("Acquisition History of Art by Tate")
```



Based on these histograms it seems like for point 1 we are spot on that the initial donations were from Henry Tate's large art collection. For points 2 and 4 a rise in both purchases and donations during the second half of the 20th century to present coincide nicely with the rise in popularity for contemporary art. For point 3 it appears that the Duveen family's financial support was not focused on buying art (it looks like there was already plenty of art) and rather more likely on the upkeep of the gallery and event programming.

If anyone actually knows the history of the Tate museum, I am very interested in confirmation of these observations as it looks pretty convincing!

Mediums of artistic expression over time.

The `medium` field is a very diverse in nature. While there is some structure to the descriptions there just over **3400 unique descriptions** written across the various pieces.

```
length(unique(artwork$medium))
```

```
## [1] 3402
```

I thought this challenge would be impossible because of the multi-dimensionality of this field, but I saw some anchors that we can use to split our data.

- A comma (,)
- "on" and "onto"
- "with"
- "and"
- semi-colon
- period

There is also some noise here that would mess up our data.

- numbers
- Repetitive terms in the verso.
- Plural terms of the same thing.
- NA values

With this in mind lets try to clean this field and use it in our data set.

```
# Need to make this ignore case
materials<-str_split(str_to_lower(artwork$medium),", |and| on | onto |On |
Onto |with |\\.|; ")
```

```
materialsdf<-list()
```

```
# I'm using a for loop here because I'm working with a list and a vector
for(i in 1:length(materials)){
  materialsdf[[i]]<-data.frame(materials=str_to_lower(
materials[[i]]),year=artwork$year[i])
}
```

```
# Now lets make a data frame
materialsdf<-do.call(rbind,materialsdf)
```

```
# Lets clean some names up.
```

```
materialsdf$materials <- str_trim(materialsdf$materials)
#Remove Verso
materialsdf$materials<-str_remove_all(materialsdf$materials,"^\\d")
materialsdf$materials<- str_remove(materialsdf$materials,"verso: ")
#Remove numbers
materialsdf$materials<-str_to_sentence(materialsdf$materials)
```

```
glimpse(materialsdf)
```

```
## Rows: 152,081
## Columns: 2
## $ materials "Watercolour", "Ink", "Chalk", "Graphite", "Paper",
"Graphite", "Paper", "Graphite", "Paper", "Graphite", "Paper", "...
## $ year      NA, NA, NA, NA, NA, NA, NA, NA, NA, 1785, 1785, 1785,
NA, NA, 1826, 1826, 1826, 1826, 1826, 1826, 1826, ...
```

We have some unknown years here. Because this isn't helpful for learning about artistic mediums over time we will take them out. We also have some empty character values in the `materials` field which we will take out as well.

To get a sense of popularity over time, we will also bin our data by century. This isn't a perfect methodology because the history of trends in art is non linear and multi-dimensional, but nevertheless there is insight to find in this manner.

The data cleansing on this isn't perfect, but I feel confident that we can list the top 5 mediums of expression from the 16th Century to the 21st Century based on the artwork from Tate. Because of varying frequencies, I am going to relate to the data in terms of proportion of all mediums employed in a given century.

```
# Make this into a tibble because the cool kids do it and filter the data
groupedMaterials<-as_tibble(materialsdf) %>%
  filter(!is.na(year), !is.na(materials), materials!="") %>%
  count(material=materials,
         century = paste0(floor(year*0.01)*100, "s")) %>%
  arrange(century,
         -n)

groupedData<-by(groupedMaterials, groupedMaterials["century"], tibble)

centuryProp<- lapply(groupedData, function(x) x$n/sum(x$n))

# I'm using for loops because its easier to read. I'm reusing the same name to
save memory!

for (i in 1:length(groupedData)){
  groupedData[[i]]<- tibble(groupedData[[i]], CenturyProp=centuryProp[[i]])
}

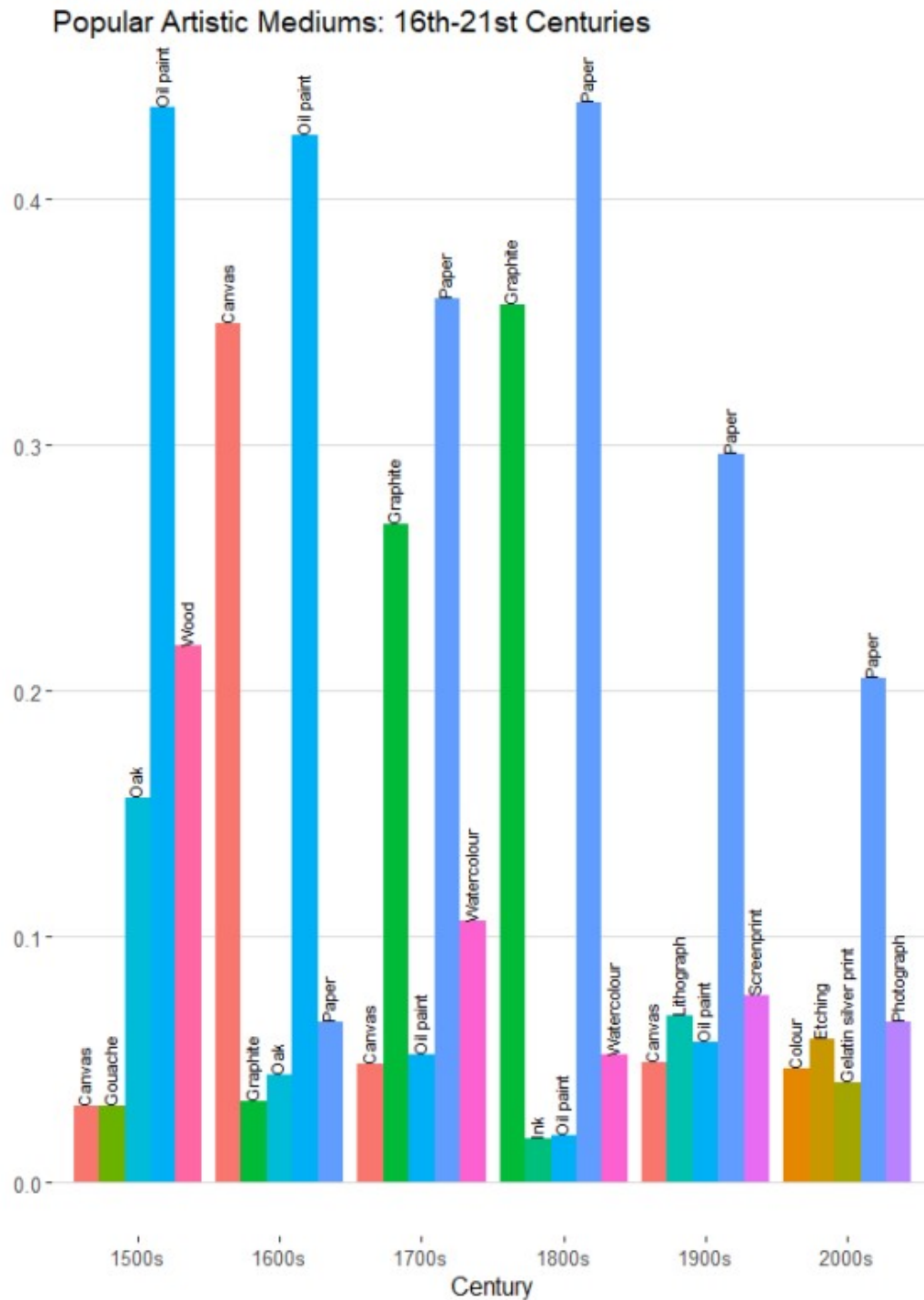
# Repeating myself, but it works.
groupedData<-do.call(rbind, groupedData)
groupedData<- by(groupedData, groupedData["century"], head, n=5)
groupedData<-do.call(rbind, groupedData)

glimpse(groupedData)

## Rows: 30
## Columns: 4
## $ material      "Oil paint", "Wood", "Oak", "Canvas", "Gouache", "Oil
paint", "Canvas", "Paper", "Oak", "Graphite", "Paper", "Grap...
## $ century       "1500s", "1500s", "1500s", "1500s", "1500s", "1600s",
"1600s", "1600s", "1600s", "1600s", "1700s", "1700s", "1700s...
## $ n             14, 7, 5, 1, 1, 78, 64, 12, 8, 6, 3070, 2283, 909, 445, 414,
32285, 26215, 3814, 1387, 1325, 12881, 3306, 2938, 24...
## $ CenturyProp   0.43750000, 0.21875000, 0.15625000, 0.03125000, 0.03125000,
0.42622951, 0.34972678, 0.06557377, 0.04371585, 0.0327...
```

Now that we have our data grouped by the top 5 mediums for each century. Lets visualize our data:

```
ggplot(data=groupedData, mapping= aes(x=century, y=CenturyProp, fill=material))+
  theme_hc()+
  geom_bar(stat="identity", position="dodge")+
  geom_text(mapping=aes(label=material), position=position_dodge(width=0.9),
vjust=0.25, hjust=0, angle=90, size=3)+
  theme(legend.position = "none")+
  labs(x="Century", y="")+
  ggtitle("Popular Artistic Mediums: 16th-21st Centuries")
```



From this visualization we can communicate the following:

- The popular medium of expression during the 16th and 17th centuries was oil paint. Wood and oak was more popular during the 16th century than Canvas. However, once canvas rose to popularity in the 17th century, wood was no longer as popular of a surface to paint with. After looking this up on the web I found out that this claim is supported [for Northern Europe](#)— which works well because all the painting available from this time period in Tate are from British artists.
- In the 18th and 19th centuries, there was a shift in popularity from using oil paint to using graphite as a medium of expression. Watercolor painting had its popularity during this time as well— interestingly enough this coincides with [the English School](#) which is when watercolor painting gained popularity in England.
- From the 18th century to present, paper has been the most popular artistic medium for art. With so much which can be done with it in the artistic space, this shouldn't be a surprise and confirms what even the less experienced with art probably already know.

Wow! looking at this data really gives us a feel for the history by seeing the various artistic mediums over

time. These insights confirm the related history of British art over the given time periods as well.

Gender breakdown over time

To visualize the gender breakdown of artists over the centuries, we need to do similar transformations of our data that we did for the artistic mediums over time.

We will be using the inner joined data of the `artists` and `artwork` tables for the visualization as the number of pieces produced by a given artist is in the `artwork` table and the gender information is in the `artists` table.

```
genderTable<- artDataIJ %>%
  count(gender,
        century=paste0(floor(year*0.01)*100,"s")) %>%
  filter(century!="NAs")

#Break down by

genderTable<-by(genderTable, genderTable["century"],tibble)

#get proportions
CentProps<- lapply(genderTable,function(x) x$n/sum(x$n))

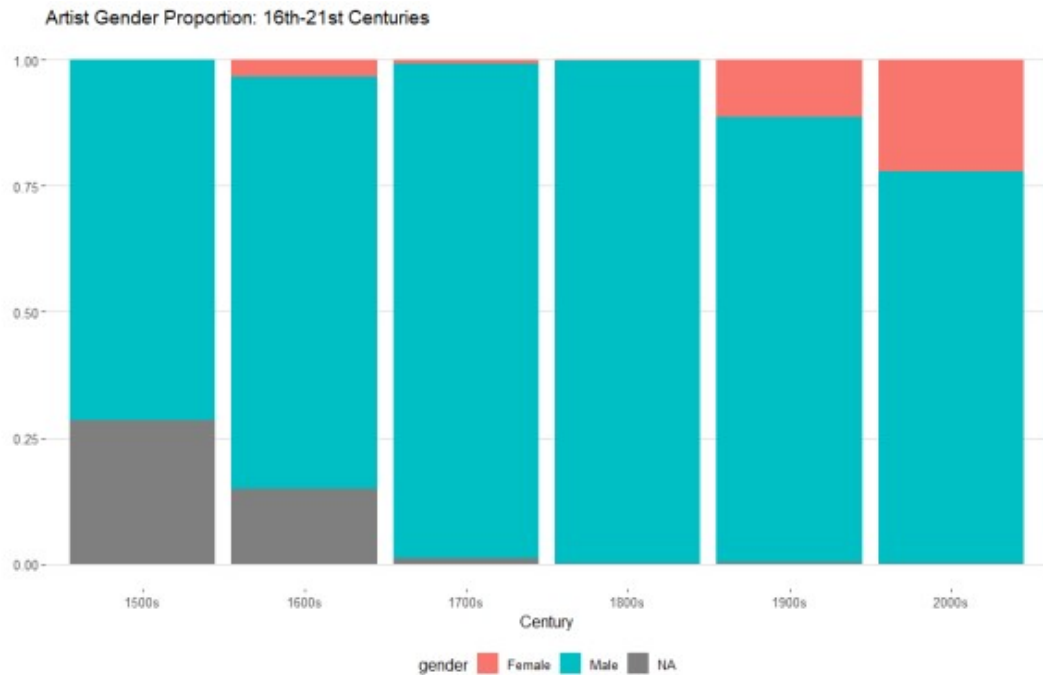
for (i in 1:length(genderTable)){
  genderTable[[i]]<- tibble(genderTable[[i]],CenturyProp=CentProps[[i]])
}

# Rebind the data

genderTable<-do.call(rbind, genderTable)
```

Once our data is transformed we can visualize the gender proportion of the artists in our data set over time.

```
ggplot(genderTable,mapping=aes(x=century, y=CenturyProp,fill=gender))+
  theme_hc()+
  geom_bar(stat="identity")+
  labs(y="",x="Century")+
  ggtitle("Artist Gender Proportion: 16th-21st Centuries")
```



What we can communicate from this visualization is:

- The gender breakdown of artists over time is predominantly male. There are no records in the data set of artwork from female artists in the 16th century and there is only a small proportion of artwork from 17th-19th centuries whose artists are female. It could be hypothesized that some of the NA values from the 16th-19th centuries were female, but there is no evidence to support this claim from data as it is presented.
- As expected with the modern social trends and revolutions, there has been a rise in artwork from female artists emerging and being featured in galleries like Tate.

Conclusion

There is a lot that we can learn about history (and in our case- art) by looking at data. What was really amazing about this project was the ability to uncover the rich history present in this data set.