

# Museum Data Join

2022-02-24

```
library(tidyverse)
```

```
Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'  
had status 1
```

```
library(tidymodels)  
library(knitr)  
library(janitor) # for the clean_names function  
library(reshape)  
library(ggplot2)  
library(dplyr)
```

## Data prep

```
# load gardner  
gardner <- read_csv(here::here("Chapter1", "data/gardner.csv")) %>%  
  janitor::clean_names()  
  
# load janson  
janson <- read_csv(here::here("Chapter1", "data/janson.csv")) %>%  
  janitor::clean_names()  
  
# load MoMA  
moma <- read_csv(here::here("Chapter1", "data/moma.csv")) %>%  
  janitor::clean_names() %>%  
  filter(year != "MISSING YEAR") %>%  
  mutate(year = as.numeric(year))
```

Warning: One or more parsing issues, see `problems()` for details

```
# load CountMoMAExh
#countmoma <- read_csv(here::here("Chapter1", "data/countmoma.csv")) %>%
# janitor::clean_names()

# load Whitney
whitney <- read_csv(here::here("Chapter1", "data/whitney.csv")) %>%
  janitor::clean_names()
```

**NOTE:** Moma data has some (very few) rows where the year is MISSING YEAR, check if those can be filled in.

Join gardner and janson with column = book, janson, gardner

```
janson <- janson %>%
  mutate(
    artist_nationality_other =
      case_when(
        artist_nationality == "French" ~ "French",
        artist_nationality == "British" ~ "British",
        artist_nationality == "American" ~ "American",
        artist_nationality == "Spanish" ~ "Spanish",
        artist_nationality == "German" ~ "German",
        TRUE ~ "Other"
      )
  ) %>%
  select(
    "artist_name",
    "edition_number",
    "year",
    "artist_nationality",
    "artist_nationality_other",
    "artist_gender",
    "artist_race",
    "artist_ethnicity",
    "book",
    "space_ratio_per_page"
  )
```

```
gardner <- gardner %>%
  mutate(
```

```

artist_nationality_other =
  case_when(
    artist_nationality == "French" ~ "French",
    artist_nationality == "British" ~ "British",
    artist_nationality == "American" ~ "American",
    artist_nationality == "Spanish" ~ "Spanish",
    artist_nationality == "German" ~ "German",
    TRUE ~ "Other"
  )
) %>%
select(
  "artist_name",
  "edition_number",
  "year",
  "artist_nationality",
  "artist_nationality_other",
  "artist_gender",
  "artist_race",
  "artist_ethnicity",
  "book",
  "space_ratio_per_page"
)

```

```

gardnerjanson <- gardner %>%
  bind_rows(janson)

sumgj <- gardnerjanson %>%
  group_by(artist_name, year)%>%
  summarize(space_ratio_per_page_total = sum(space_ratio_per_page))

```

`summarise()` has grouped output by 'artist\_name'. You can override using the `.groups` argument.

```

gardnerjanson <- gardnerjanson%>%
  left_join(sumgj)%>%
  select(-space_ratio_per_page)%>%
  unique()

```

Joining, by = c("artist\_name", "year")

```
gardnerjanson <- gardnerjanson %>%
  group_by(artist_name)%>%
  mutate(artist_unique_id = cur_group_id())
```

---

Create moma with complete years

```
moma_complete_years <- moma %>%
  count(artist_name, year)%>%
  mutate(count = 1)%>%
  arrange(artist_name, year)%>%
  group_by(artist_name)%>%
  ungroup() %>%
  complete(artist_name, year = c(min(year):max(year))) %>%
  mutate(
    moma_count = if_else(is.na(count), 0, count)
  ) %>%
  arrange(artist_name, year)%>%
  group_by(artist_name)%>%
  mutate(moma_count_to_date = cumsum(moma_count))%>%
  ungroup()%>%
  select(artist_name, year, moma_count, moma_count_to_date)
```

Create whitney with complete years

```
whitney_complete_years <- whitney %>%
  count(artist_name, year)%>%
  mutate(count = 1)%>%
  arrange(artist_name, year)%>%
  group_by(artist_name)%>%
  ungroup() %>%
  complete(artist_name, year = c(min(year):max(year))) %>%
  mutate(
    whitney_count = if_else(is.na(count), 0, count)
  ) %>%
  arrange(artist_name, year)%>%
  group_by(artist_name)%>%
  mutate(whitney_count_to_date = cumsum(whitney_count))%>%
  ungroup()%>%
  select(artist_name, year, whitney_count, whitney_count_to_date)
```

```
gardnerjanson_museums <- gardnerjanson %>%
  left_join(moma_complete_years, by = c("artist_name", "year")) %>%
  left_join(whitney_complete_years, by = c("artist_name", "year"))
```

```
gardnerjanson_museums[is.na(gardnerjanson_museums)] <- 0
```

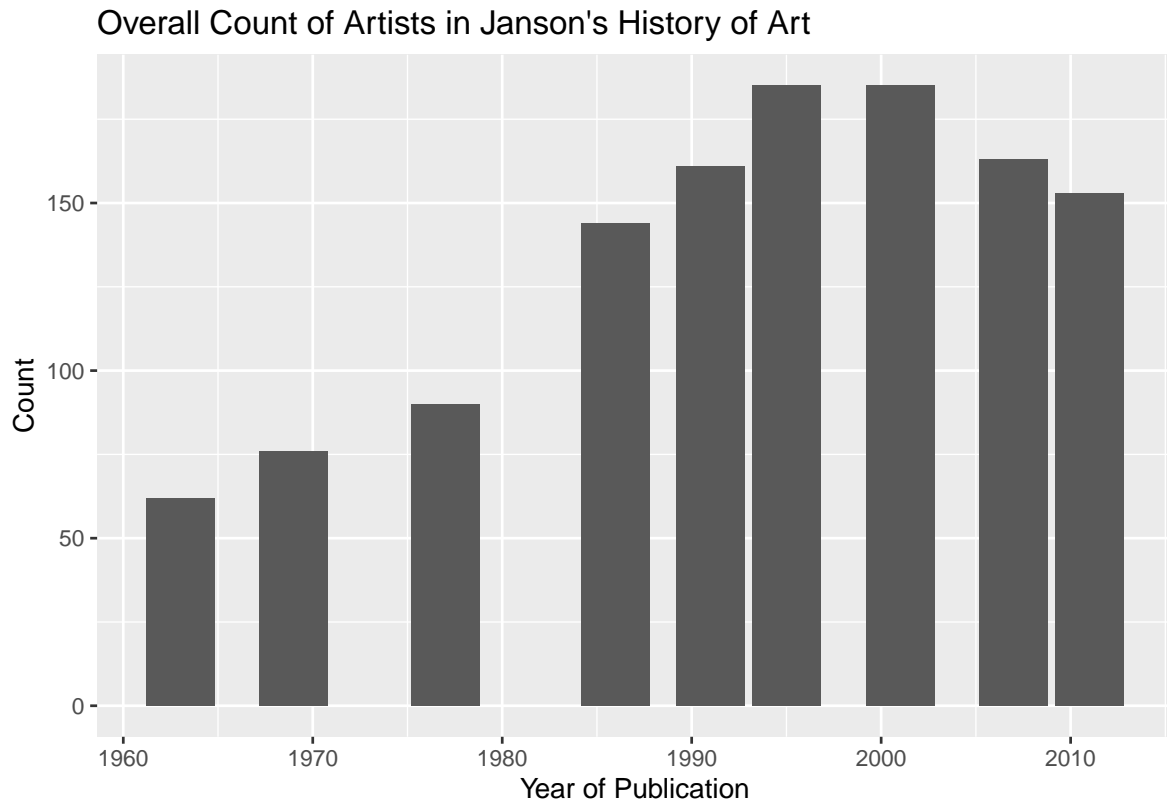
potentially needed for heat map:

```
#gardnerjanson_museums <- gardnerjanson_museums %>%
  # mutate(moma_cat = case_when(moma_count_to_date >= 1 ~ '1',
  #   TRUE ~ '0'))
```

```
#gardnerjanson_museums <- gardnerjanson_museums %>%
  # mutate(whitney_cat = case_when(whitney_count_to_date >= 1 ~ '1',
  #   TRUE ~ '0'))
```

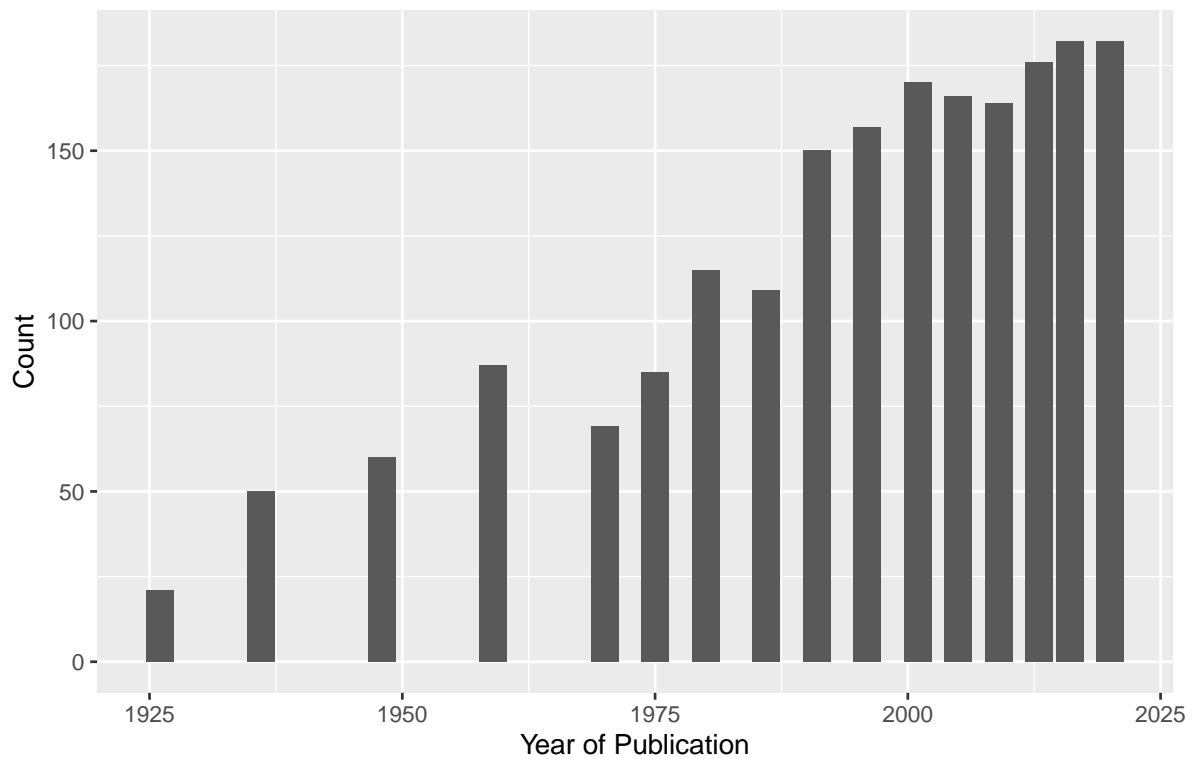
## Exploratory data analysis

```
ggplot(gardnerjanson %>% filter(book == "janson"), aes(x = year))+
  geom_bar()+
  labs(title = "Overall Count of Artists in Janson's History of Art",
    x = "Year of Publication",
    y = "Count")
```

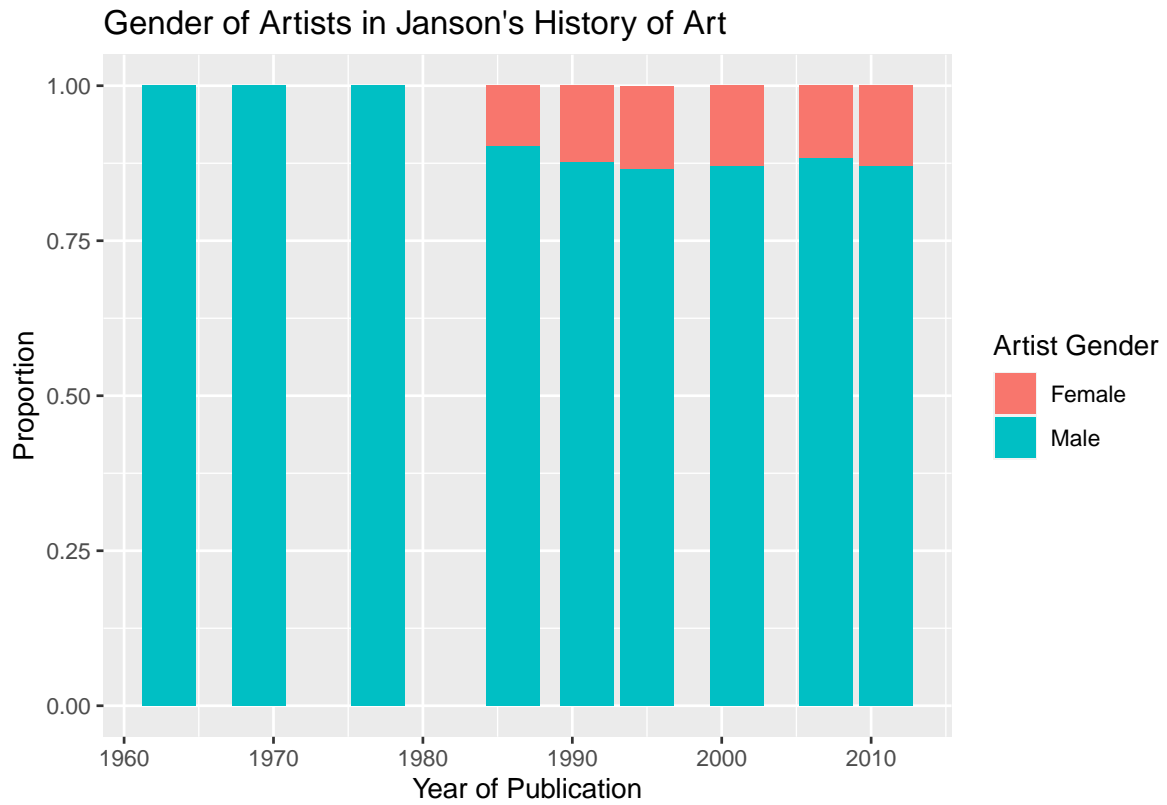


```
ggplot(gardnerjanson %>% filter(book == "gardner"), aes(x = year))+  
  geom_bar()+  
  labs(title = "Overall Count of Artists in Gardner's Art Through the Ages",  
        x = "Year of Publication",  
        y = "Count")
```

Overall Count of Artists in Gardner's Art Through the Ages

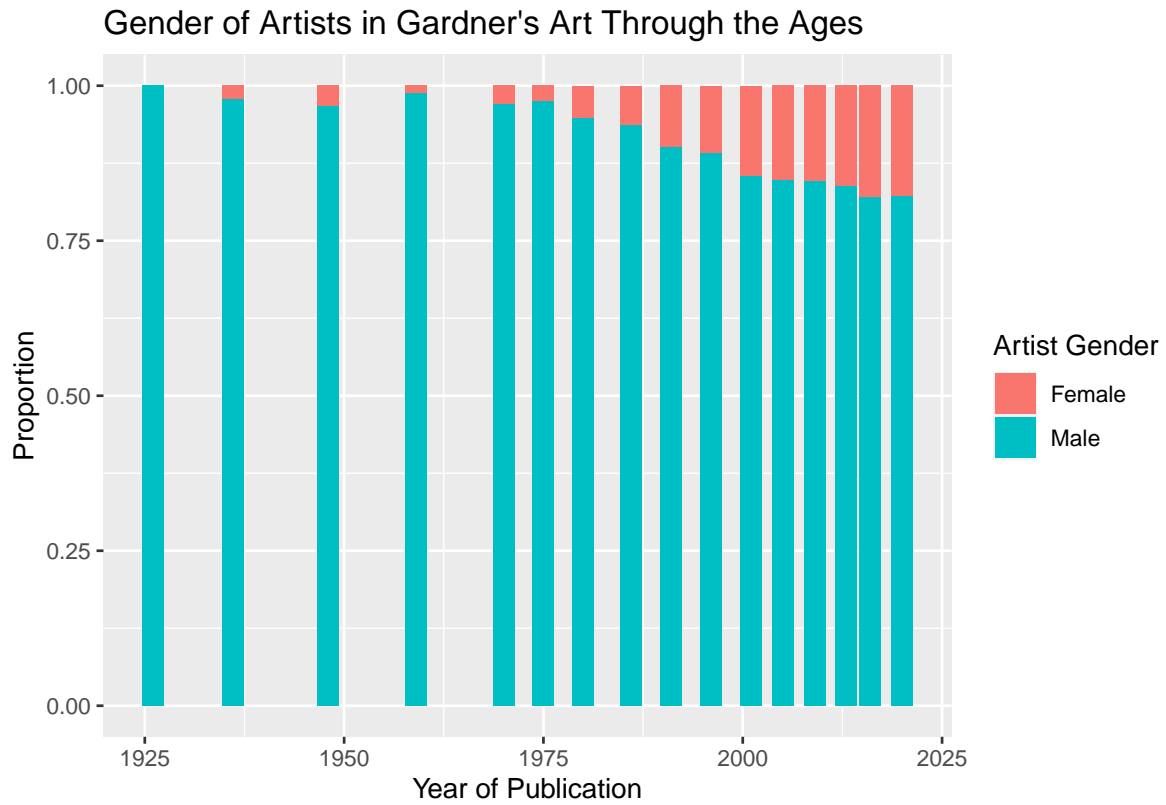


```
ggplot(gardnerjanson %>% filter(book == "janson"), aes(x = year, fill = artist_gender))+
  geom_bar(position = "fill")+
  labs(title = "Gender of Artists in Janson's History of Art",
        x = "Year of Publication",
        y = "Proportion",
        fill = "Artist Gender")
```

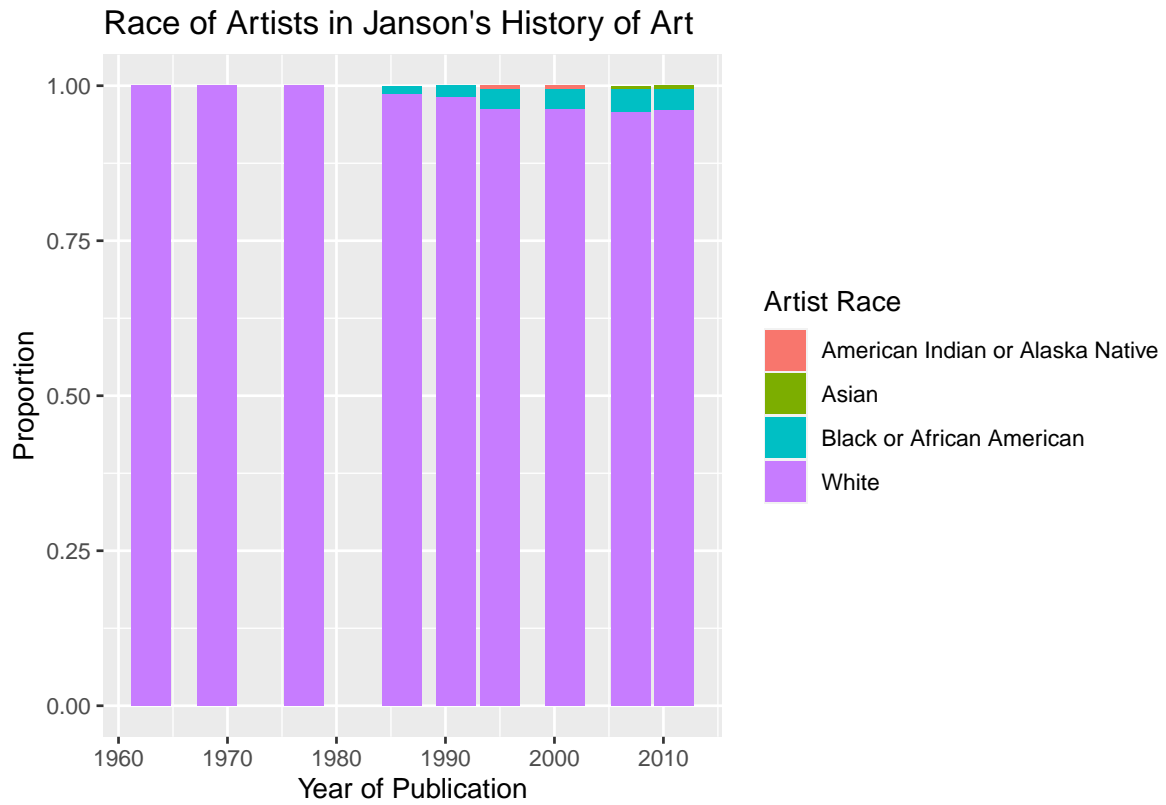


```
ggplot(gardnerjanson %>% filter(book == "gardner", artist_gender != "N/A"), aes(x = year, fi.  
  geom_bar(position = "fill")+  
  labs(title = "Gender of Artists in Gardner's Art Through the Ages",  
        x = "Year of Publication",  
        y = "Proportion",  
        fill = "Artist Gender")
```

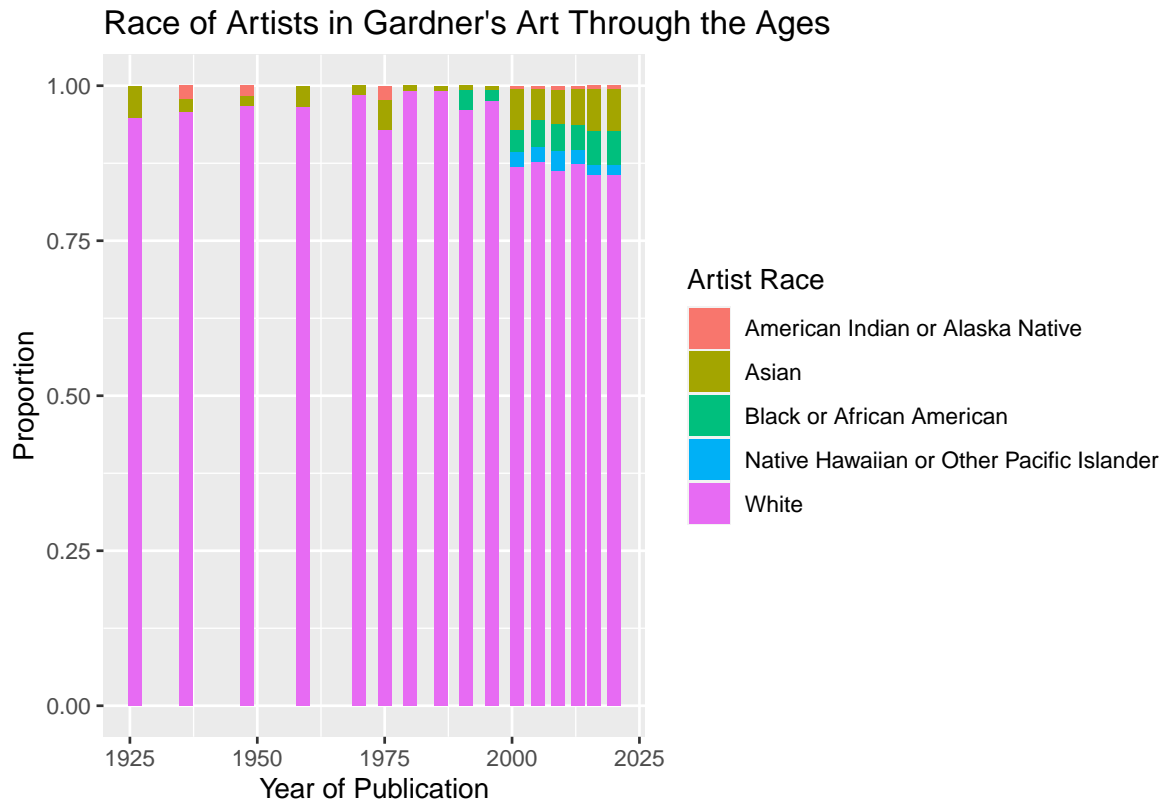




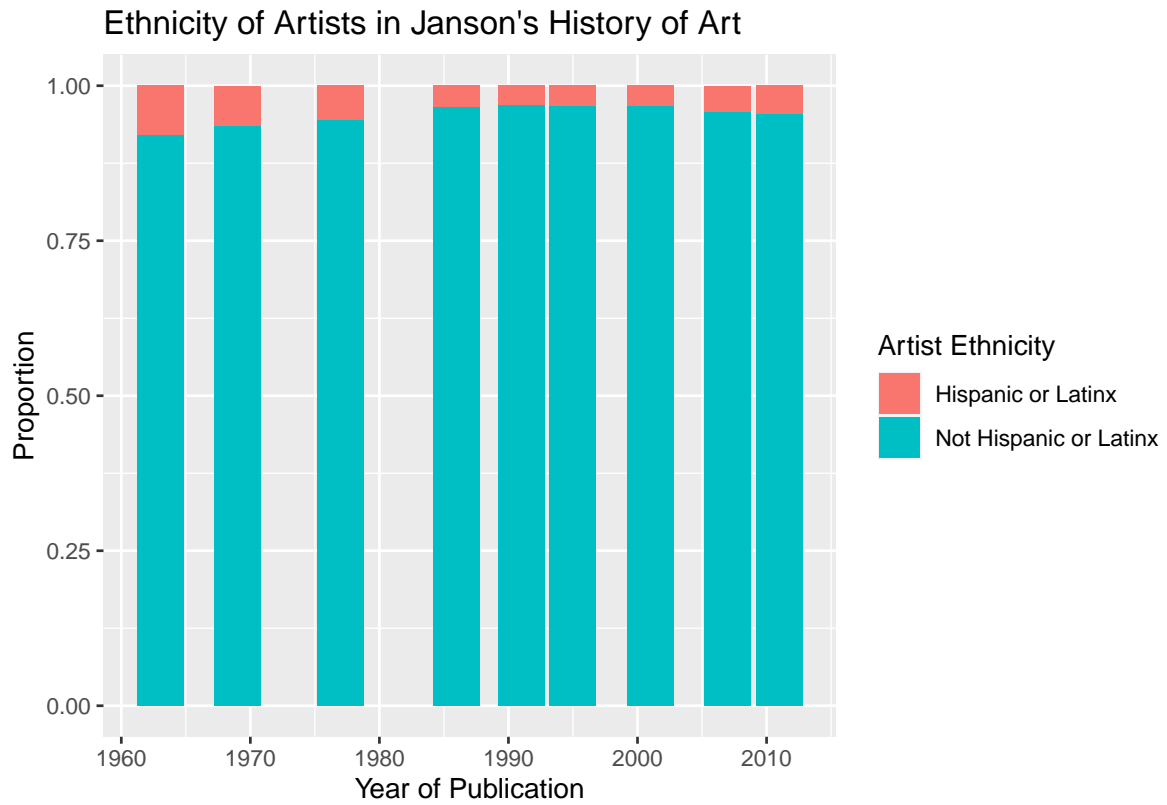
```
ggplot(gardnerjanson %>% filter(book == "janson"), aes(x = year, fill = artist_race))+
  geom_bar(position = "fill")+
  labs(title = "Race of Artists in Janson's History of Art",
       x = "Year of Publication",
       y = "Proportion",
       fill = "Artist Race")
```



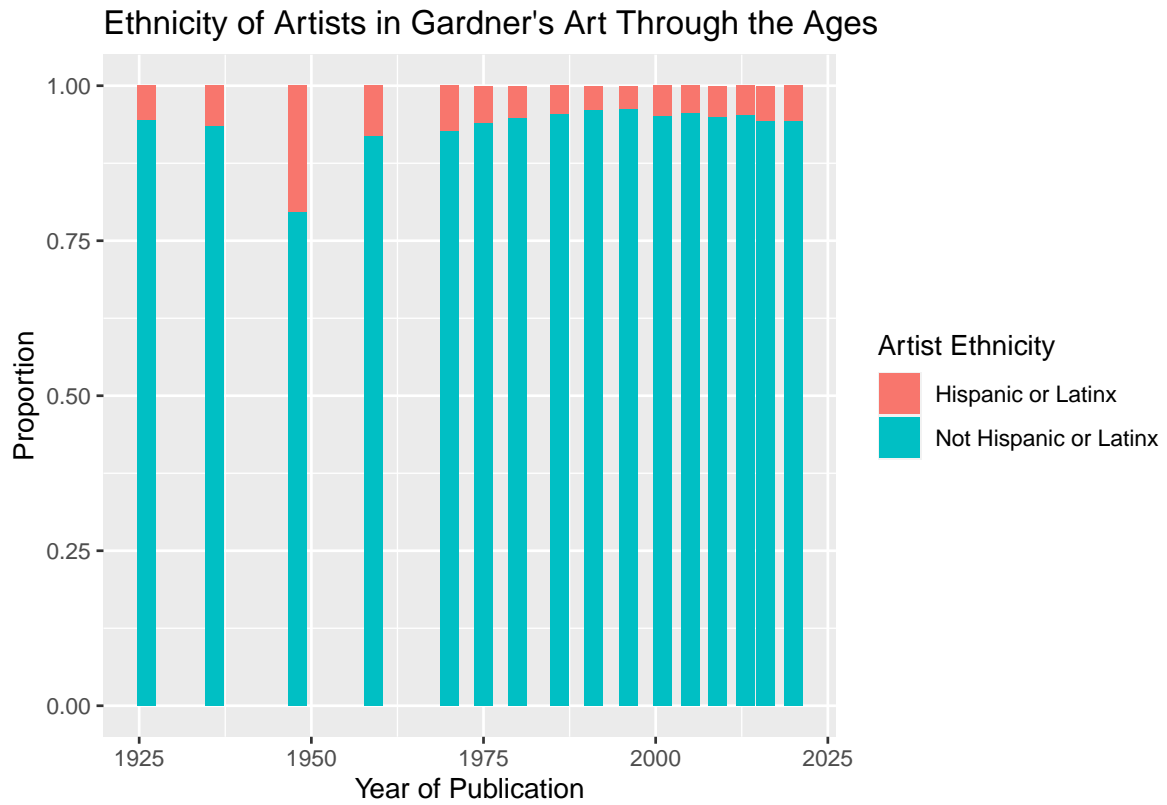
```
ggplot(gardnerjanson %>% filter(book == "gardner", artist_race != "N/A"), aes(x = year, fill = "Artist Race"))  
  geom_bar(position = "fill") +  
  labs(title = "Race of Artists in Gardner's Art Through the Ages",  
        x = "Year of Publication",  
        y = "Proportion",  
        fill = "Artist Race")
```



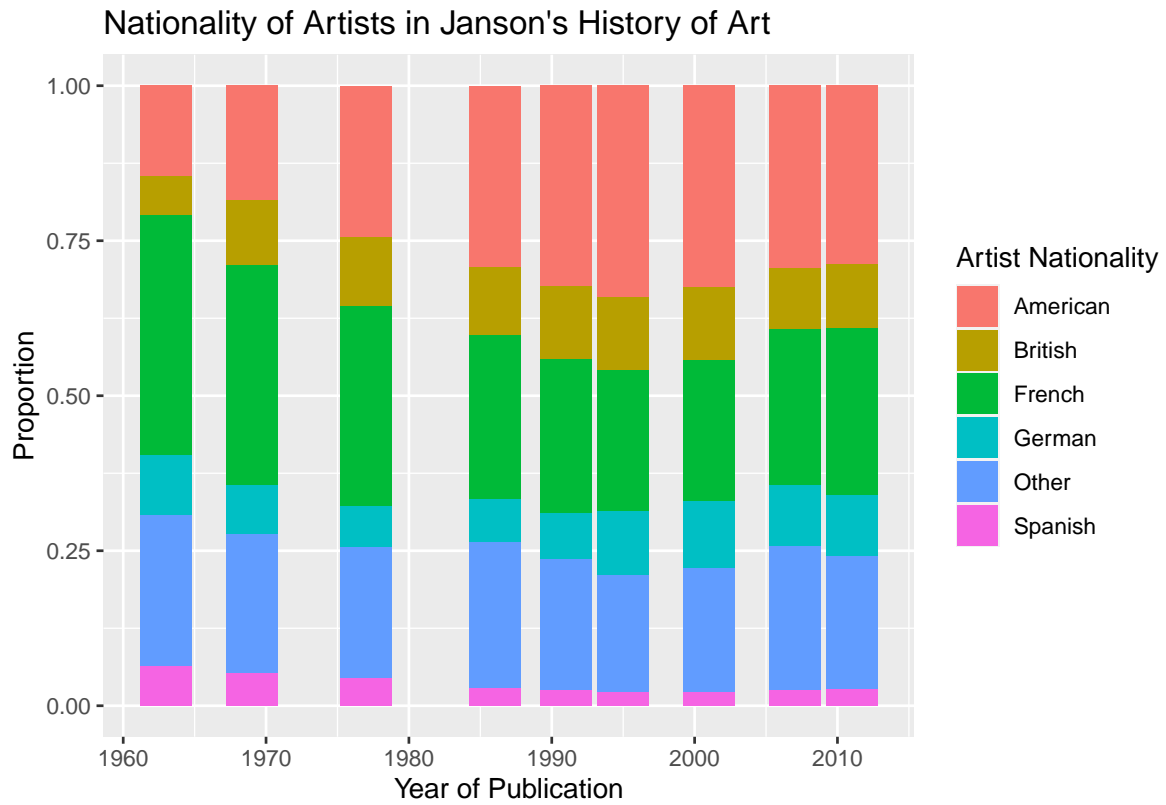
```
ggplot(gardnerjanson %>% filter(book == "janson"), aes(x = year, fill = artist_ethnicity))+
  geom_bar(position = "fill")+
  labs(title = "Ethnicity of Artists in Janson's History of Art",
       x = "Year of Publication",
       y = "Proportion",
       fill = "Artist Ethnicity")
```



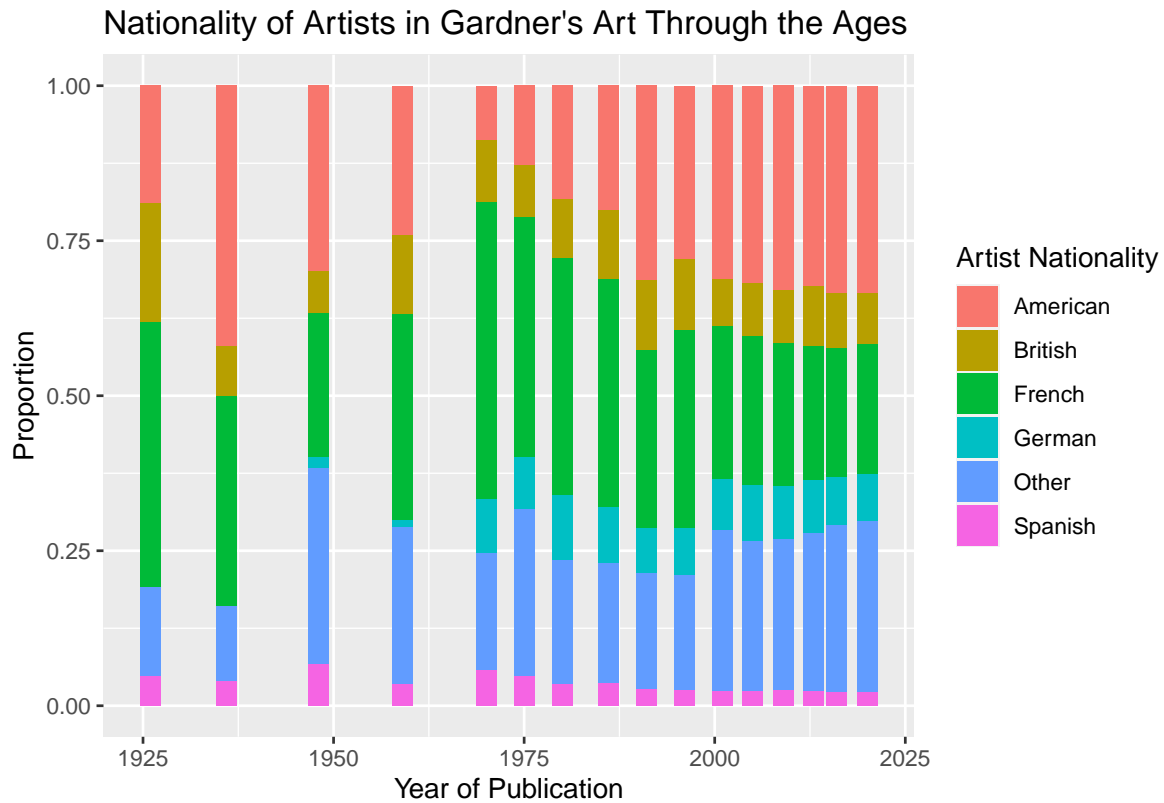
```
ggplot(gardnerjanson %>% filter(book == "gardner", artist_ethnicity != "N/A"), aes(x = year,  
  geom_bar(position = "fill")+  
  labs(title = "Ethnicity of Artists in Gardner's Art Through the Ages",  
    x = "Year of Publication",  
    y = "Proportion",  
    fill = "Artist Ethnicity"))
```



```
ggplot(gardnerjanson %>% filter(book == "janson"), aes(x = year, fill = artist_nationality_of)) +
  geom_bar(position = "fill") +
  labs(title = "Nationality of Artists in Janson's History of Art",
       x = "Year of Publication",
       y = "Proportion",
       fill = "Artist Nationality")
```



```
ggplot(gardnerjanson %>% filter(book == "gardner", artist_nationality_other != "N/A"), aes(x = "Year of Publication", y = "Proportion", fill = "Artist Nationality"))
  geom_bar(position = "fill")
  labs(title = "Nationality of Artists in Gardner's Art Through the Ages",
        x = "Year of Publication",
        y = "Proportion",
        fill = "Artist Nationality")
```

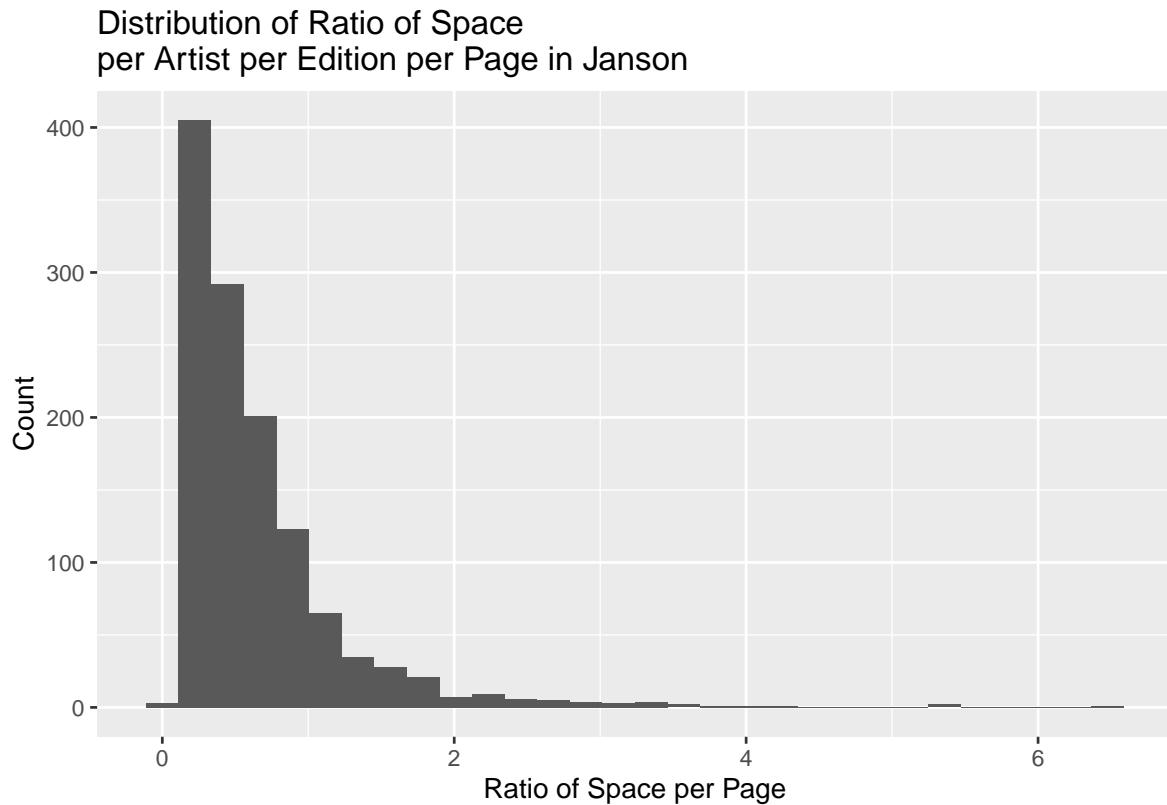


## Bivariate Plots and MLR with `space_ratio_per_page_total`

Janson:

```
ggplot(gardnerjanson_museums %>% filter(book == "janson"),
       aes(x = space_ratio_per_page_total, fill = edition_number)) +
  geom_histogram() +
  labs(
    title = "Distribution of Ratio of Space\nper Artist per Edition per Page in Janson",
    x = "Ratio of Space per Page",
    y = "Count"
  )
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



```
summary(janson$space_ratio_per_page)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0946	0.2531	0.3387	0.3967	0.4855	1.2785

```
median <- median(janson$space_ratio_per_page)
```

The visualization above is right skewed, unimodal and assymetrical. Therefore, we would want to look at the median and IQR. The median of `space_ratio_per_page_total` in Janson is 0.3387305.

`space_ratio_per_page` with Gender

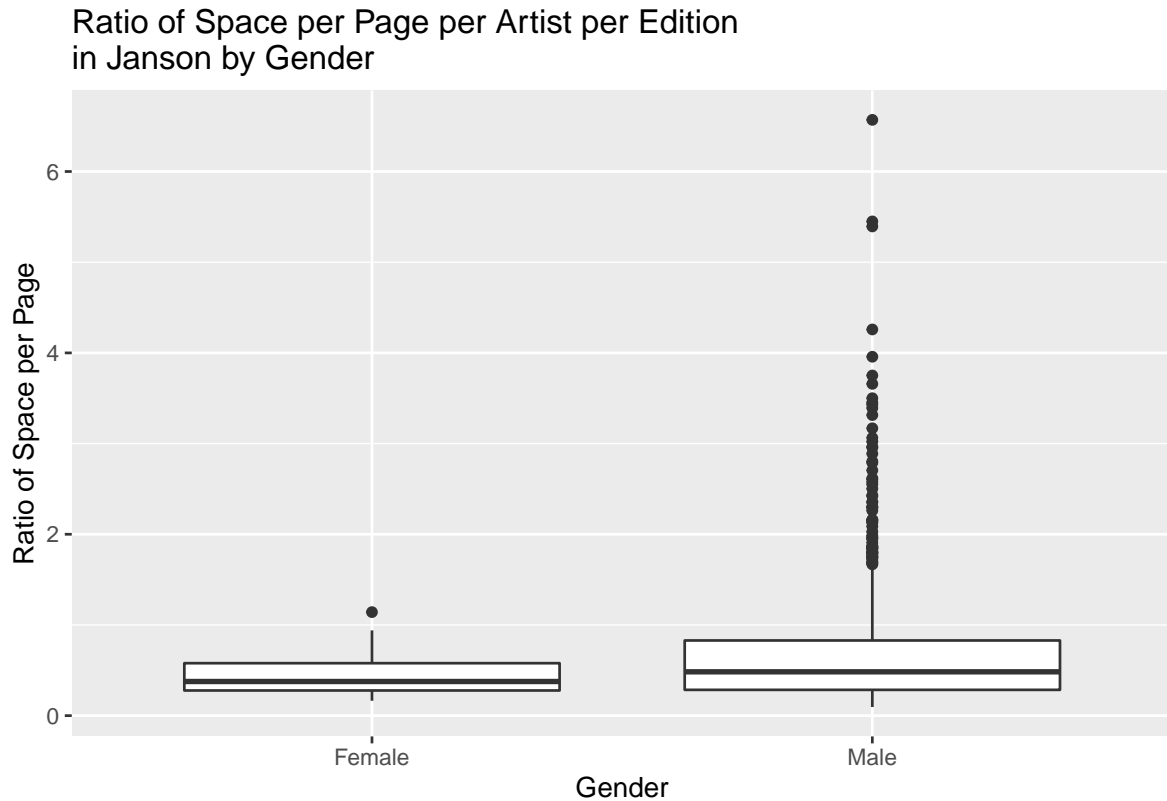
```
ggplot(gardnerjanson_museums %>% filter(book == "janson"),
       aes(x = artist_gender, y = space_ratio_per_page_total)) +
  geom_boxplot() +
  labs(
```



```

title = "Ratio of Space per Page per Artist per Edition\nin Janson by Gender",
x = "Gender",
y = "Ratio of Space per Page"
)

```



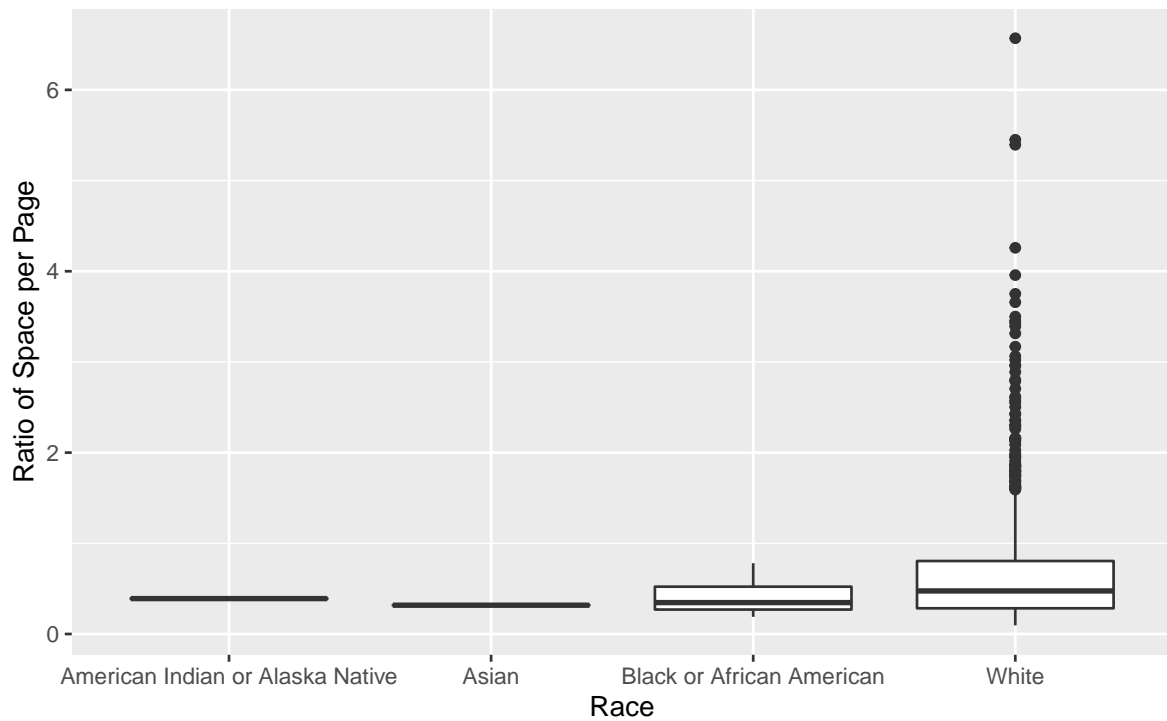
space\_ratio\_per\_page with Race

```

ggplot(gardnerjanson_museums %>% filter(book == "janson"),
       aes(x = artist_race, y = space_ratio_per_page_total)) +
  geom_boxplot() +
  labs(
    title = "Ratio of Space per Page per Artist per Edition\nin Janson by Race",
    x = "Race",
    y = "Ratio of Space per Page"
  )

```

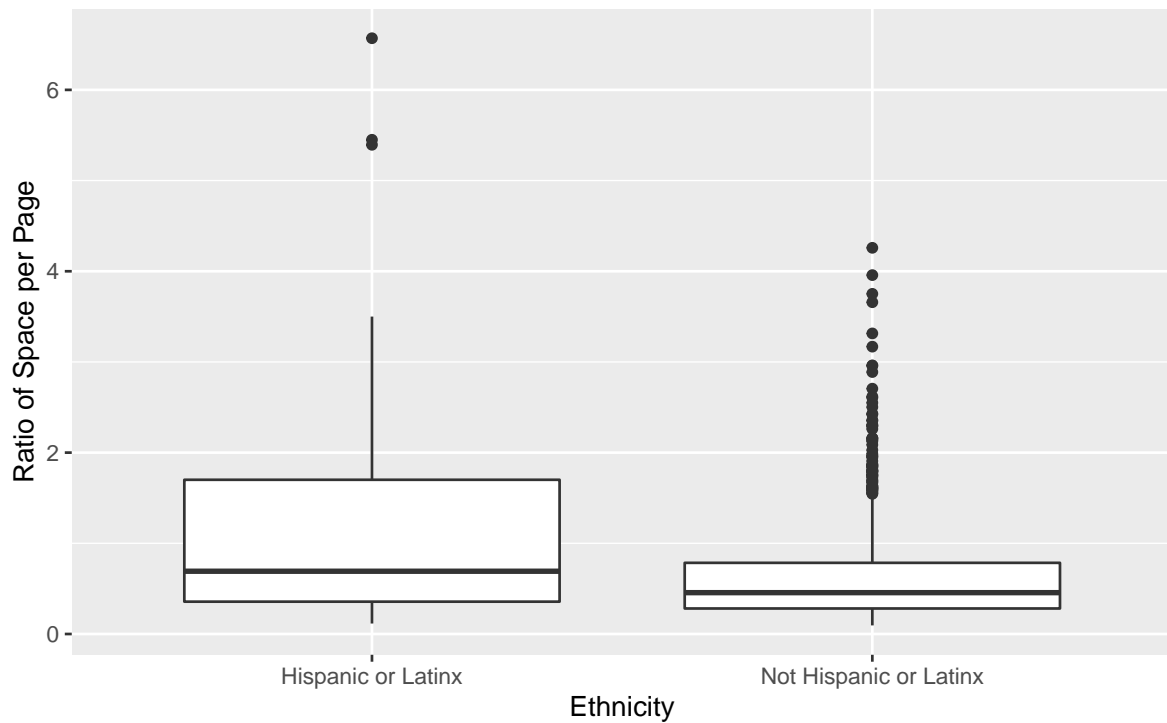
Ratio of Space per Page per Artist per Edition  
in Janson by Race



space\_ratio\_per\_page with Ethnicity

```
ggplot(gardnerjanson_museums %>% filter(book == "janson"),
       aes(x = artist_ethnicity, y = space_ratio_per_page_total)) +
  geom_boxplot() +
  labs(
    title = "Ratio of Space per Page per Artist per Edition\nin Janson by Ethnicity",
    x = "Ethnicity",
    y = "Ratio of Space per Page"
  )
```

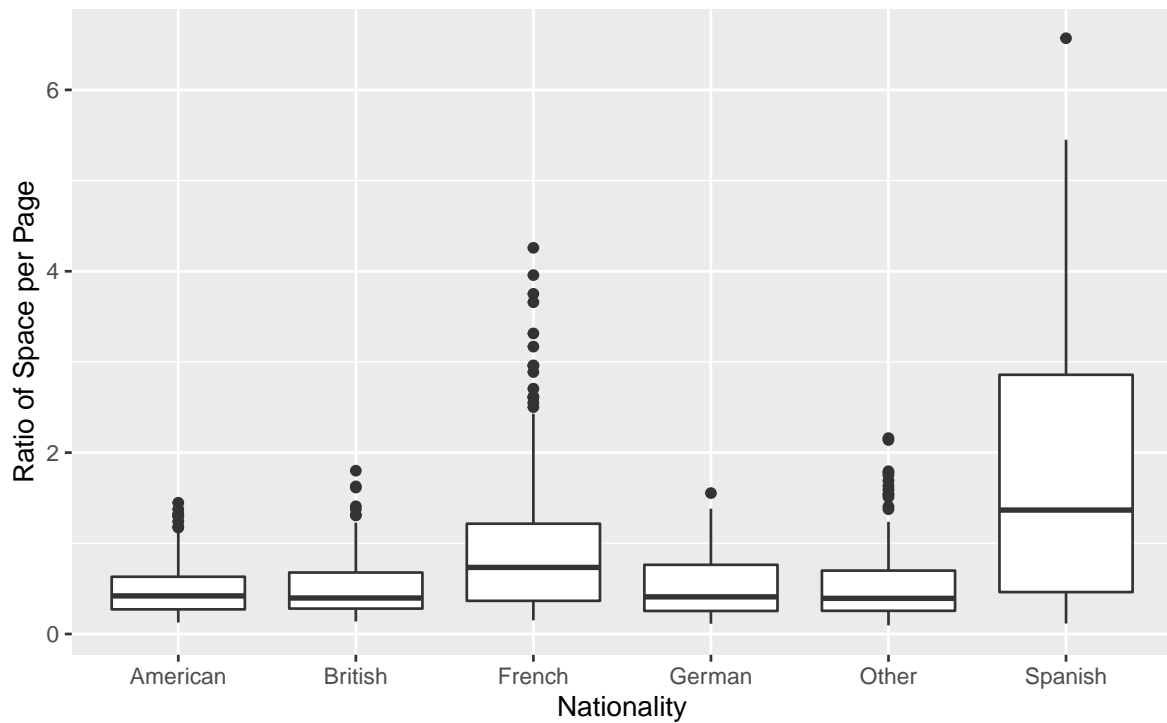
Ratio of Space per Page per Artist per Edition  
in Janson by Ethnicity



space\_ratio\_per\_page with Nationality

```
ggplot(gardnerjanson_museums %>% filter(book == "janson"),
       aes(x = artist_nationality_other, y = space_ratio_per_page_total)) +
  geom_boxplot() +
  labs(
    title = "Ratio of Space per Page per Artist per Edition\nin Janson by Nationality",
    x = "Nationality",
    y = "Ratio of Space per Page"
  )
```

Ratio of Space per Page per Artist per Edition  
in Janson by Nationality



MLR of Janson with Gender + Race + Nationality + Ethnicity ~ space\_ratio\_per\_page

```
jansonMLR <- lm(space_ratio_per_page_total ~ artist_race +
  artist_ethnicity +
  artist_gender +
  artist_nationality_other +
  moma_count_to_date +
  whitney_count_to_date,
  data = gardnerjanson_museums %>% filter(book == "janson")
)
summary(jansonMLR)
```

Call:

```
lm(formula = space_ratio_per_page_total ~ artist_race + artist_ethnicity +
  artist_gender + artist_nationality_other + moma_count_to_date +
  whitney_count_to_date, data = gardnerjanson_museums %>% filter(book ==
  "janson"))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.5866	-0.2831	-0.0976	0.1931	4.6319

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.999e-01	4.078e-01	0.981	0.327
artist_raceAsian	-1.798e-01	5.434e-01	-0.331	0.741
artist_raceBlack or African American	-9.906e-02	3.988e-01	-0.248	0.804
artist_raceWhite	-2.281e-02	3.853e-01	-0.059	0.953
artist_ethnicityNot Hispanic or Latinx	-9.541e-03	1.440e-01	-0.066	0.947
artist_genderMale	1.069e-01	5.400e-02	1.980	0.048
artist_nationality_otherBritish	3.281e-02	5.877e-02	0.558	0.577
artist_nationality_otherFrench	4.108e-01	4.659e-02	8.817	< 2e-16
artist_nationality_otherGerman	-8.681e-03	6.059e-02	-0.143	0.886
artist_nationality_otherOther	-3.172e-05	4.855e-02	-0.001	0.999
artist_nationality_otherSpanish	1.171e+00	1.750e-01	6.692	3.37e-11
moma_count_to_date	9.425e-03	1.994e-03	4.727	2.54e-06
whitney_count_to_date	2.071e-04	3.723e-03	0.056	0.956

(Intercept)	
artist_raceAsian	
artist_raceBlack or African American	
artist_raceWhite	
artist_ethnicityNot Hispanic or Latinx	
artist_genderMale	*
artist_nationality_otherBritish	
artist_nationality_otherFrench	***
artist_nationality_otherGerman	
artist_nationality_otherOther	
artist_nationality_otherSpanish	***
moma_count_to_date	***
whitney_count_to_date	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5395 on 1206 degrees of freedom

Multiple R-squared: 0.2149, Adjusted R-squared: 0.2071

F-statistic: 27.51 on 12 and 1206 DF, p-value: < 2.2e-16

```
glance(jansonMLR)$r.squared
```

```
[1] 0.2148817
```

```
glance(jansonMLR)$adj.r.squared
```

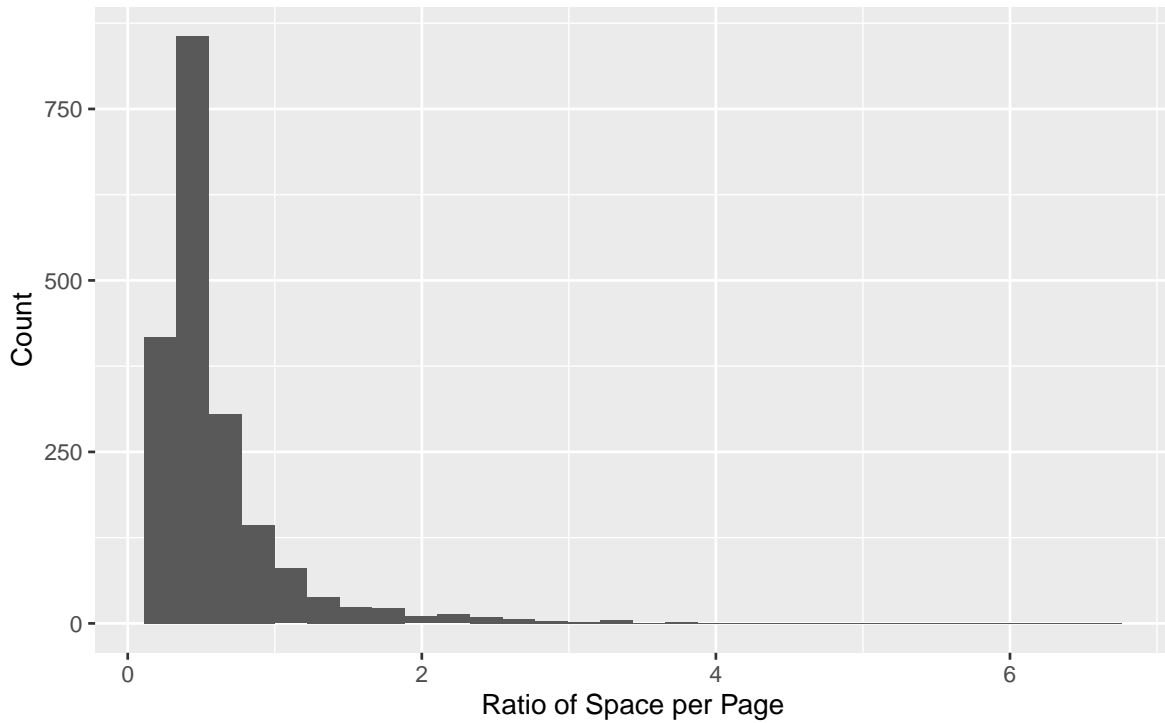
```
[1] 0.2070696
```

Gardner:

```
ggplot(  
  gardnerjanson_museums %>% filter(book == "gardner"),  
  aes(x = space_ratio_per_page_total)  
) +  
  geom_histogram() +  
  labs(  
    title = "Distribution of Ratio of Space\nper Artist per Edition per Page in Gardner",  
    x = "Ratio of Space per Page",  
    y = "Count"  
  )
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.

Distribution of Ratio of Space  
per Artist per Edition per Page in Gardner



```
gardnerjanson_museums %>%
  filter(book == "gardner") %>%
  summarize(
    min = min(space_ratio_per_page_total),
    q1 = quantile(space_ratio_per_page_total, 0.25),
    median = median(space_ratio_per_page_total),
    mean = mean(space_ratio_per_page_total),
    q3 = quantile(space_ratio_per_page_total, 0.75),
    max = max(space_ratio_per_page_total)
  )
```

# A tibble: 334 x 7

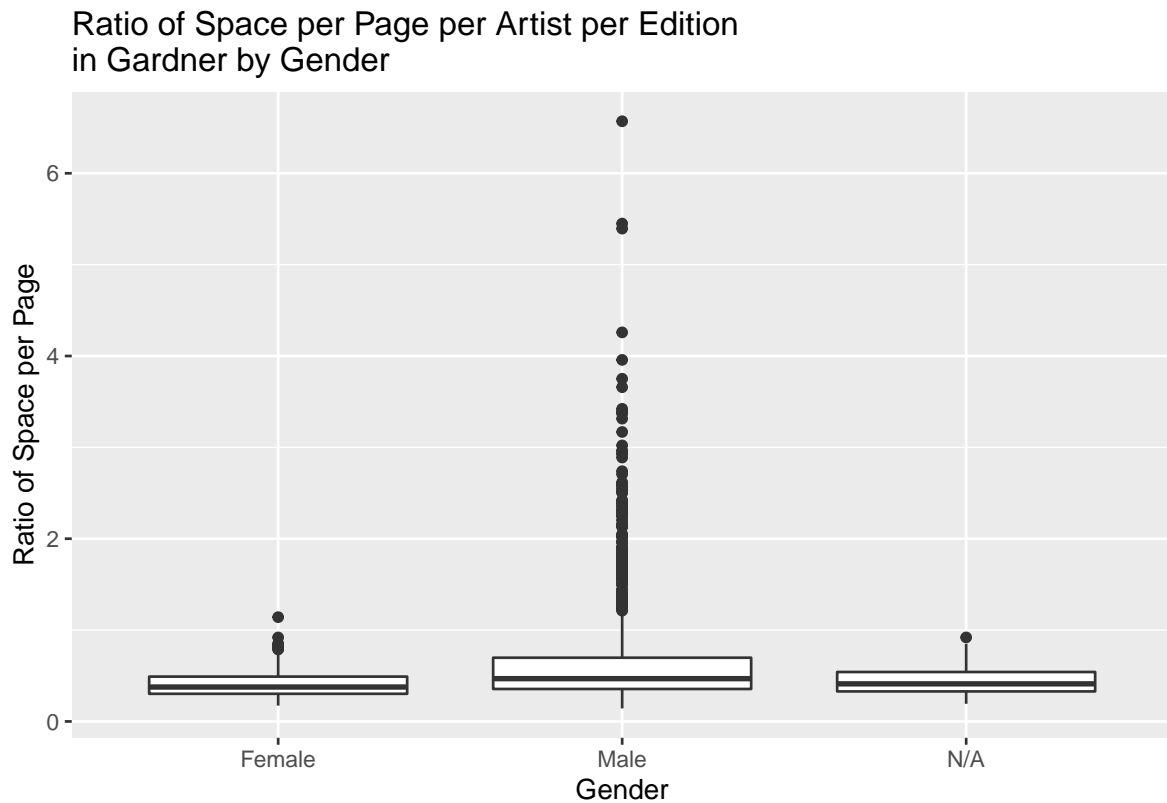
	artist_name	min	q1	median	mean	q3	max
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Aaron Douglas	0.303	0.369	0.388	0.411	0.460	0.544
2	Adélaïde Labille-Guiard	0.409	0.411	0.413	0.452	0.474	0.534
3	Adolphe William Bouguereau	0.218	0.294	0.298	0.339	0.390	0.532
4	Albert Bierstadt	0.368	0.382	0.431	0.428	0.457	0.506

5	Albert Pinkham Ryder	0.315	0.361	0.461	0.415	0.467	0.476
6	Albert Sands Southworth	0.222	0.287	0.294	0.305	0.348	0.356
7	Alexander Gardner	0.348	0.348	0.348	0.348	0.348	0.348
8	Alexandre Cabanel	0.502	0.502	0.502	0.502	0.502	0.502
9	Alfred Stieglitz	0.415	0.490	0.502	0.659	0.551	1.30
10	Alfredo Guido	0.306	0.306	0.306	0.306	0.306	0.306

# ... with 324 more rows

space\_ratio\_per\_page vs Gender

```
ggplot(gardnerjanson_museums %>% filter(book == "gardner"),
  aes(x = artist_gender, y = space_ratio_per_page_total)
) +
  geom_boxplot() +
  labs(
    title = "Ratio of Space per Page per Artist per Edition\nin Gardner by Gender",
    x = "Gender",
    y = "Ratio of Space per Page"
  )
```

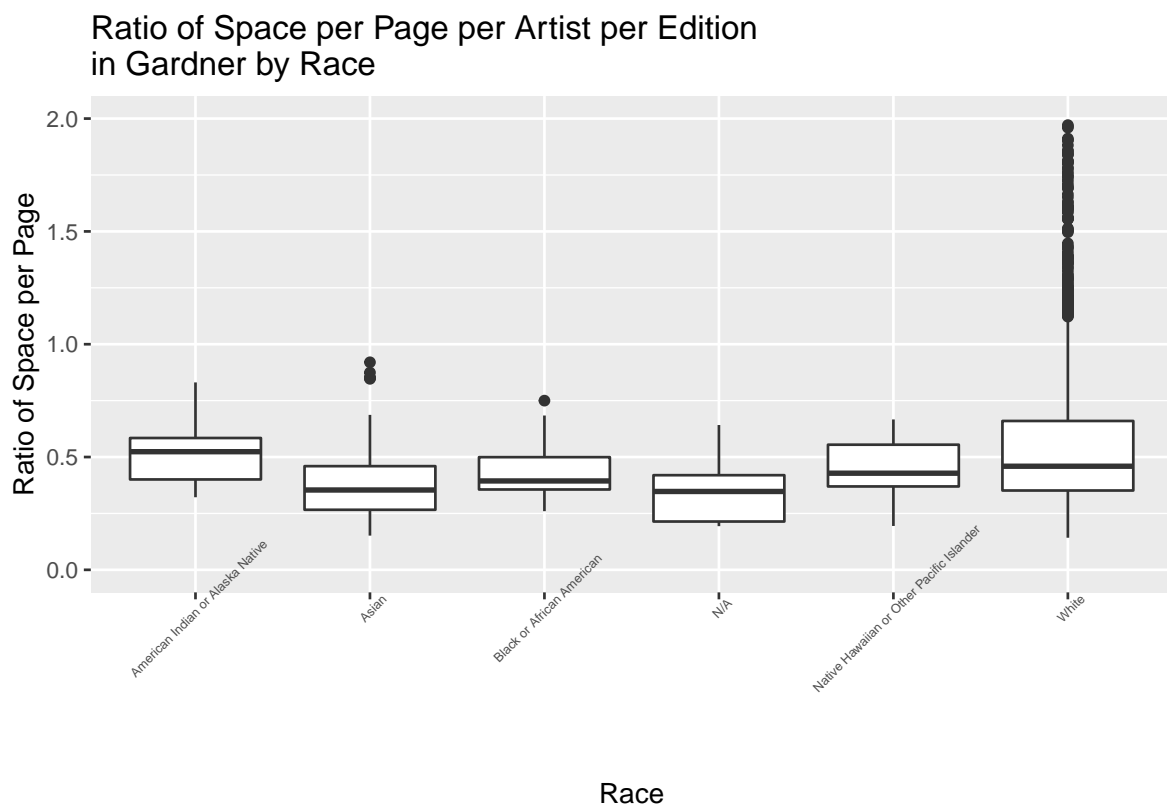




space\_ratio\_per\_page vs Race

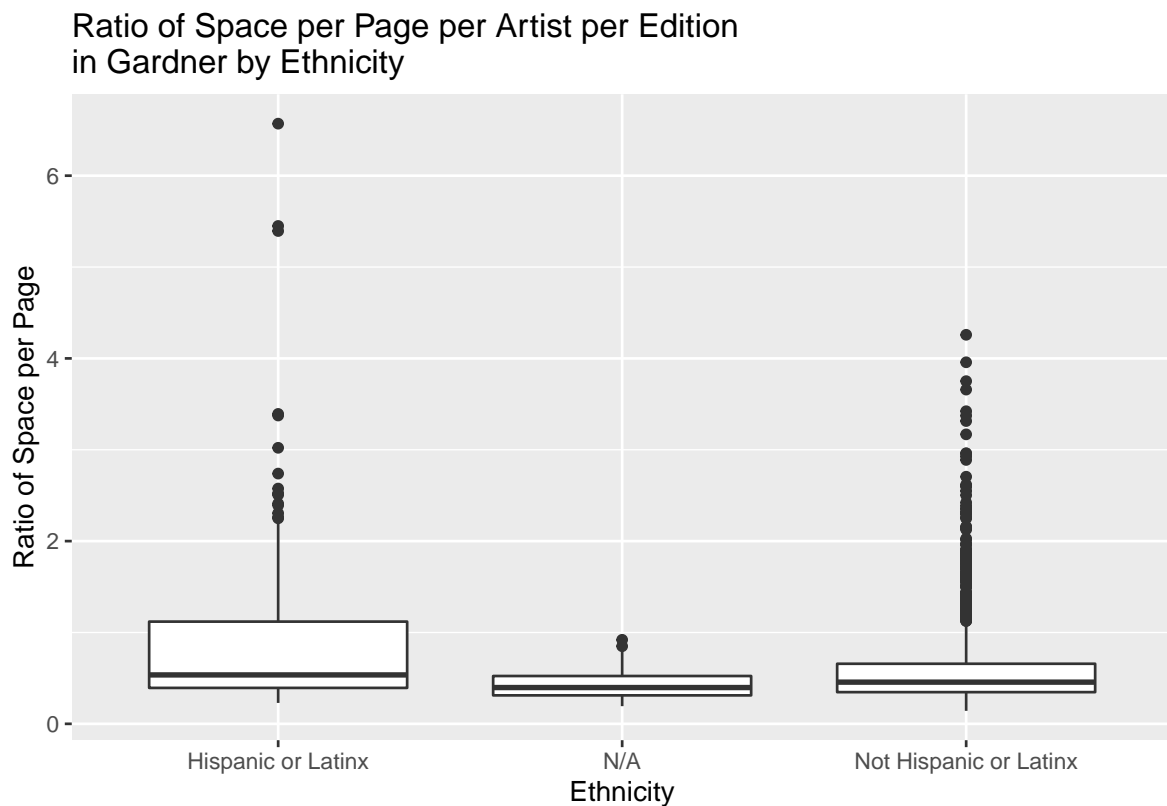
```
ggplot(  
  gardnerjanson_museums %>% filter(book == "gardner"),  
  aes(x = artist_race, y = space_ratio_per_page_total)  
) +  
  geom_boxplot() +  
  labs(  
    title = "Ratio of Space per Page per Artist per Edition\nin Gardner by Race",  
    x = "Race",  
    y = "Ratio of Space per Page"  
  ) +  
  theme(axis.text.x = element_text(angle = 45, size = 5)) +  
  ylim(0, 2.0)
```

Warning: Removed 51 rows containing non-finite values (stat\_boxplot).



space\_ratio\_per\_page vs Ethnicity

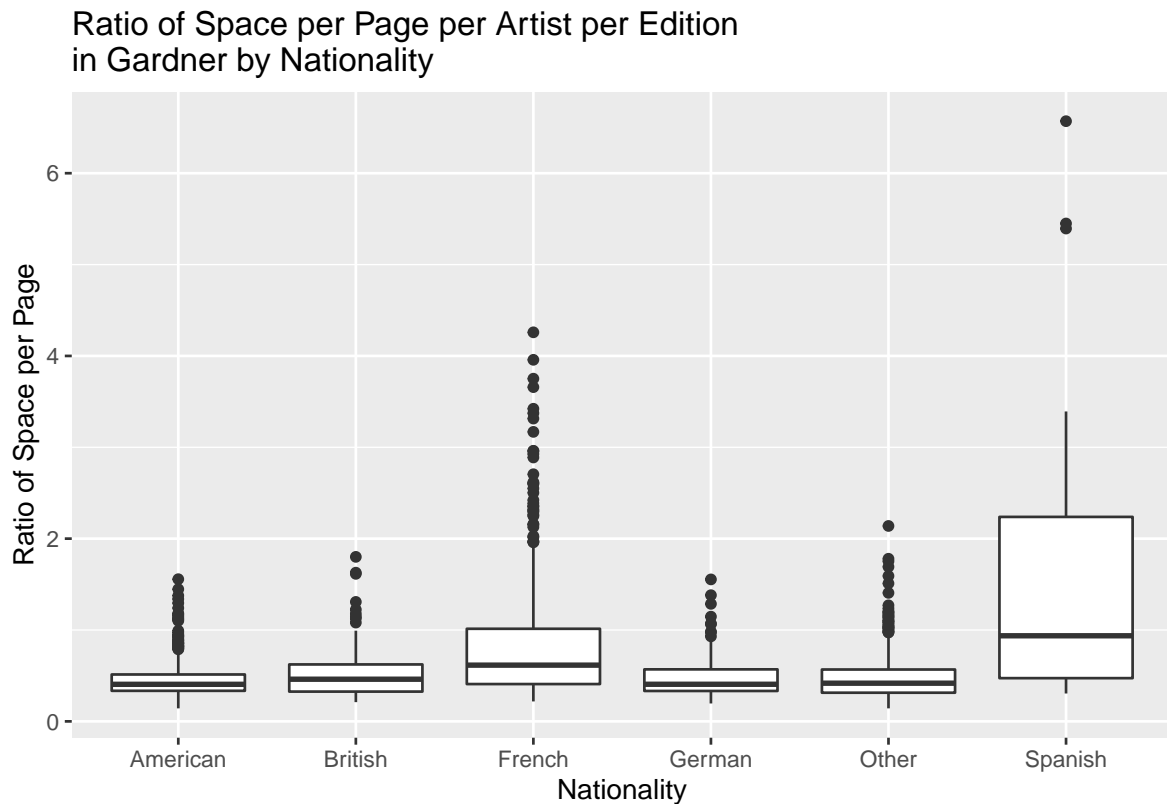
```
ggplot(
  gardnerjanson_museums %>% filter(book == "gardner"),
  aes(x = artist_ethnicity, y = space_ratio_per_page_total)
) +
  geom_boxplot() +
  labs(
    title = "Ratio of Space per Page per Artist per Edition\nin Gardner by Ethnicity",
    x = "Ethnicity",
    y = "Ratio of Space per Page"
  )
)
```



space\_ratio\_per\_page vs Nationality

```
ggplot(
  gardnerjanson_museums %>% filter(book == "gardner"),
  aes(x = artist_nationality_other, y = space_ratio_per_page_total)
) +
```

```
geom_boxplot() +
labs(
  title = "Ratio of Space per Page per Artist per Edition\nin Gardner by Nationality",
  x = "Nationality",
  y = "Ratio of Space per Page"
)
```



GardnerMLR Demographic

```
gardnerMLR <- lm(space_ratio_per_page_total ~ artist_race +
  artist_ethnicity +
  artist_gender +
  artist_nationality_other +
  moma_count_to_date +
  whitney_count_to_date,
data = gardnerjanson_museums %>% filter(book == "gardner")
)
summary(gardnerMLR)
```

Call:

```
lm(formula = space_ratio_per_page_total ~ artist_race + artist_ethnicity +  
    artist_gender + artist_nationality_other + moma_count_to_date +  
    whitney_count_to_date, data = gardnerjanson_museums %>% filter(book ==  
    "gardner"))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.2410	-0.1989	-0.0693	0.0882	4.9413

Coefficients:

	Estimate	Std. Error
(Intercept)	0.369341	0.177880
artist_raceAsian	-0.085242	0.166393
artist_raceBlack or African American	-0.007132	0.173678
artist_raceN/A	-0.172717	0.167182
artist_raceNative Hawaiian or Other Pacific Islander	-0.085124	0.180836
artist_raceWhite	0.002815	0.164019
artist_ethnicityN/A	-0.124660	0.228008
artist_ethnicityNot Hispanic or Latinx	-0.011957	0.069883
artist_genderMale	0.109756	0.033519
artist_genderN/A	0.275479	0.224553
artist_nationality_otherBritish	0.042408	0.040975
artist_nationality_otherFrench	0.331675	0.031441
artist_nationality_otherGerman	-0.021700	0.043768
artist_nationality_otherOther	0.012956	0.034829
artist_nationality_otherSpanish	0.899780	0.096468
moma_count_to_date	0.008220	0.001411
whitney_count_to_date	-0.002032	0.002120

	t value	Pr(> t )
(Intercept)	2.076	0.03799 *
artist_raceAsian	-0.512	0.60850
artist_raceBlack or African American	-0.041	0.96725
artist_raceN/A	-1.033	0.30168
artist_raceNative Hawaiian or Other Pacific Islander	-0.471	0.63789
artist_raceWhite	0.017	0.98631
artist_ethnicityN/A	-0.547	0.58463
artist_ethnicityNot Hispanic or Latinx	-0.171	0.86417
artist_genderMale	3.274	0.00108 **
artist_genderN/A	1.227	0.22005
artist_nationality_otherBritish	1.035	0.30080
artist_nationality_otherFrench	10.549	< 2e-16 ***
artist_nationality_otherGerman	-0.496	0.62009

artist_nationality_otherOther	0.372	0.70995	
artist_nationality_otherSpanish	9.327	< 2e-16	***
moma_count_to_date	5.826	6.64e-09	***
whitney_count_to_date	-0.959	0.33792	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4489 on 1926 degrees of freedom

Multiple R-squared: 0.2144, Adjusted R-squared: 0.2078

F-statistic: 32.84 on 16 and 1926 DF, p-value: < 2.2e-16

```
glance(gardnerMLR)$r.squared
```

```
[1] 0.2143531
```

```
glance(gardnerMLR)$adj.r.squared
```

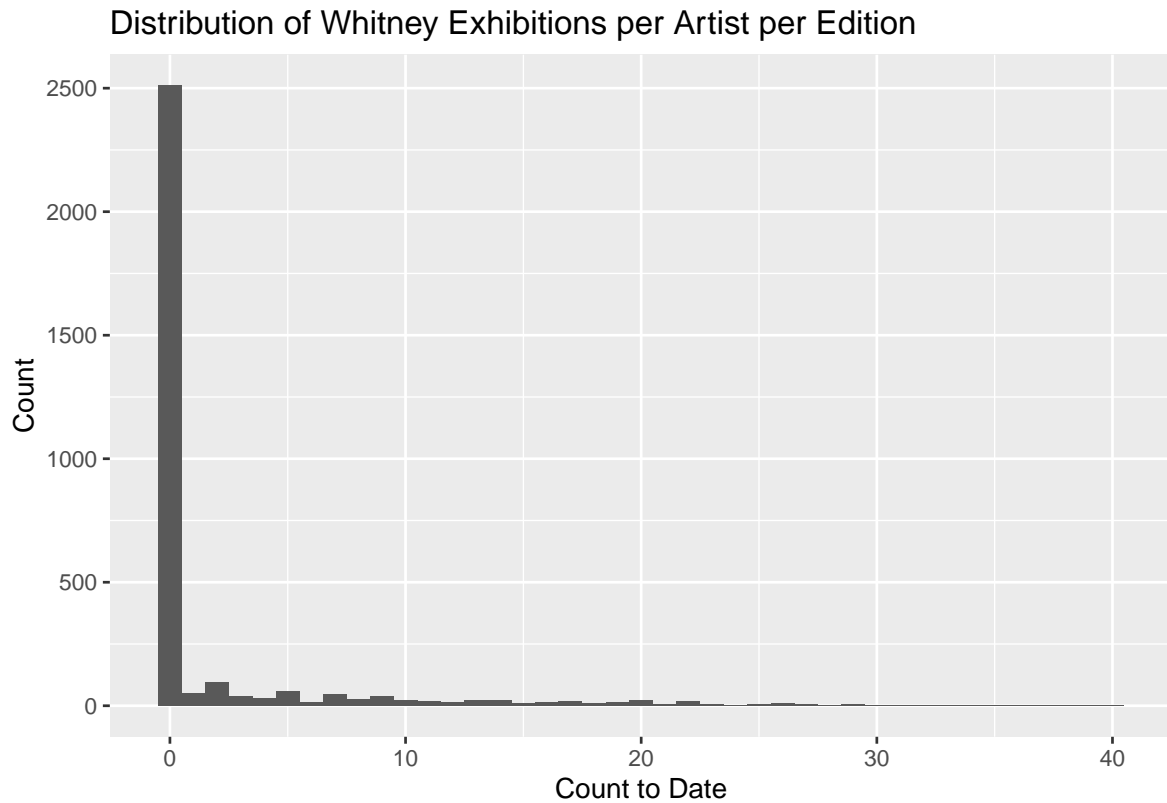
```
[1] 0.2078265
```

MLR demographic with jansongardner

## Museums

whitney eda

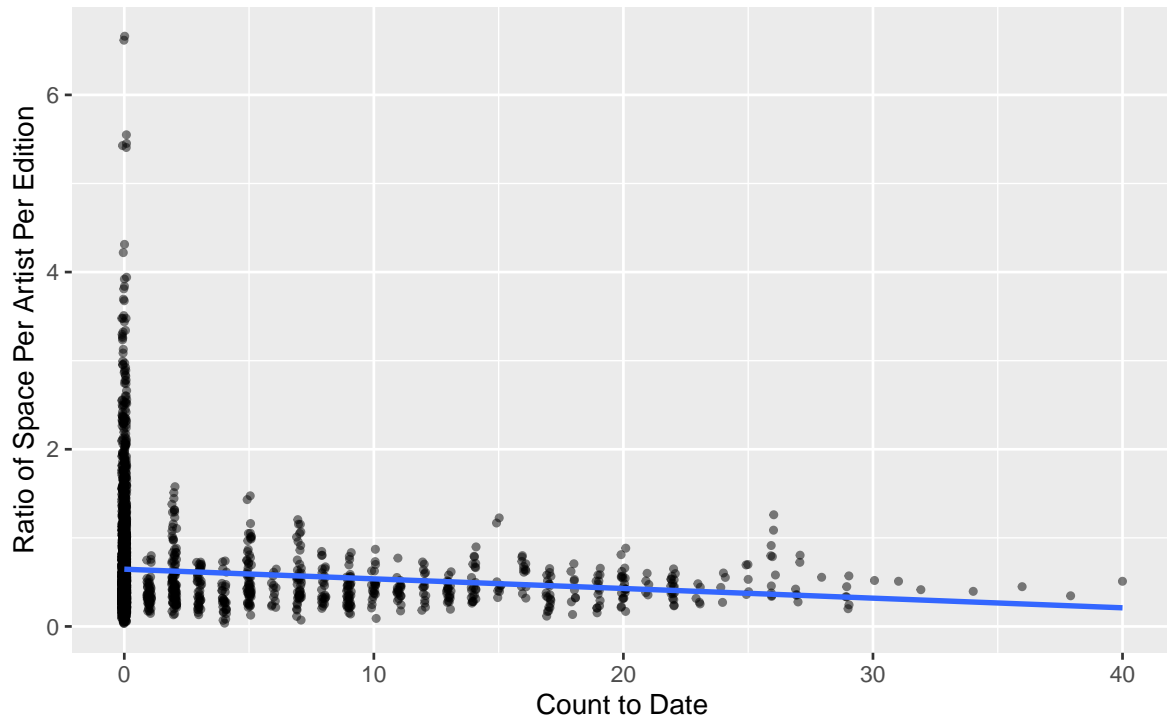
```
ggplot(gardnerjanson_museums, aes(x = whitney_count_to_date)) +
  geom_histogram(binwidth = 1) +
  labs(
    title = "Distribution of Whitney Exhibitions per Artist per Edition",
    x = "Count to Date",
    y = "Count"
  )
```



```
ggplot(gardnerjanson_museums, aes(x = whitney_count_to_date, y = space_ratio_per_page_total)) +
  geom_point(position=position_jitter(h=0.1, w=0.1), alpha = 0.5, size = 1) +
  geom_smooth(method = "lm", se = FALSE)+
  labs(
    title = "Ratio of Space Per Artist Per Edition by\nNumber of Exhibitions per Artist per Edition",
    x = "Count to Date",
    y = "Ratio of Space Per Artist Per Edition"
  )
```

`geom\_smooth()` using formula 'y ~ x'

Ratio of Space Per Artist Per Edition by  
Number of Exhibitions per Artist per Editions in the Whitney

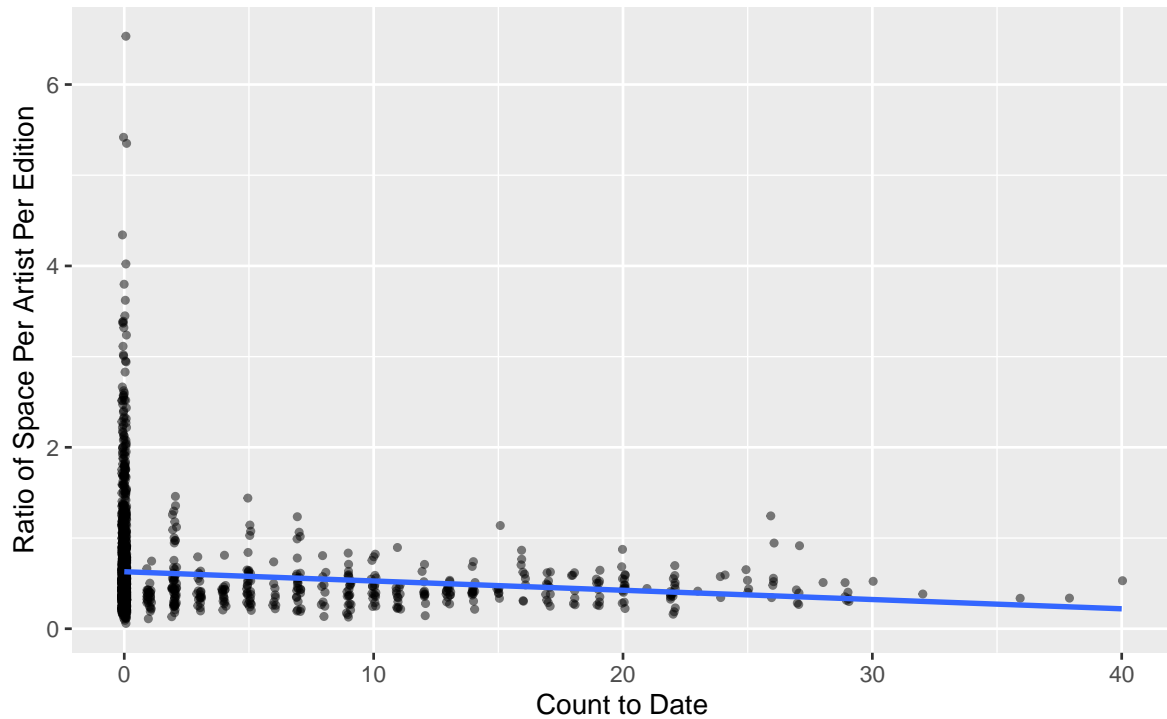


```
#corr_coef
corr_coef_whitney <- cor(gardnerjanson_museums$whitney_count_to_date,
  gardnerjanson_museums$space_ratio_per_page_total,
  use = "complete.obs")
```

```
ggplot(gardnerjanson_museums %>% filter(book == "gardner"), aes(x = whitney_count_to_date, y = space_ratio_per_page_total)) +
  geom_point(position=position_jitter(h=0.1, w=0.1), alpha = 0.5, size = 1) +
  geom_smooth(method = "lm", se = FALSE)+
  labs(
    title = "Ratio of Space Per Artist Per Edition in Gardner by\nNumber of Exhibitions per Artist",
    x = "Count to Date",
    y = "Ratio of Space Per Artist Per Edition"
  )
```

`geom\_smooth()` using formula 'y ~ x'

Ratio of Space Per Artist Per Edition in Gardner by  
Number of Exhibitions per Artist per Editions in the Whitney



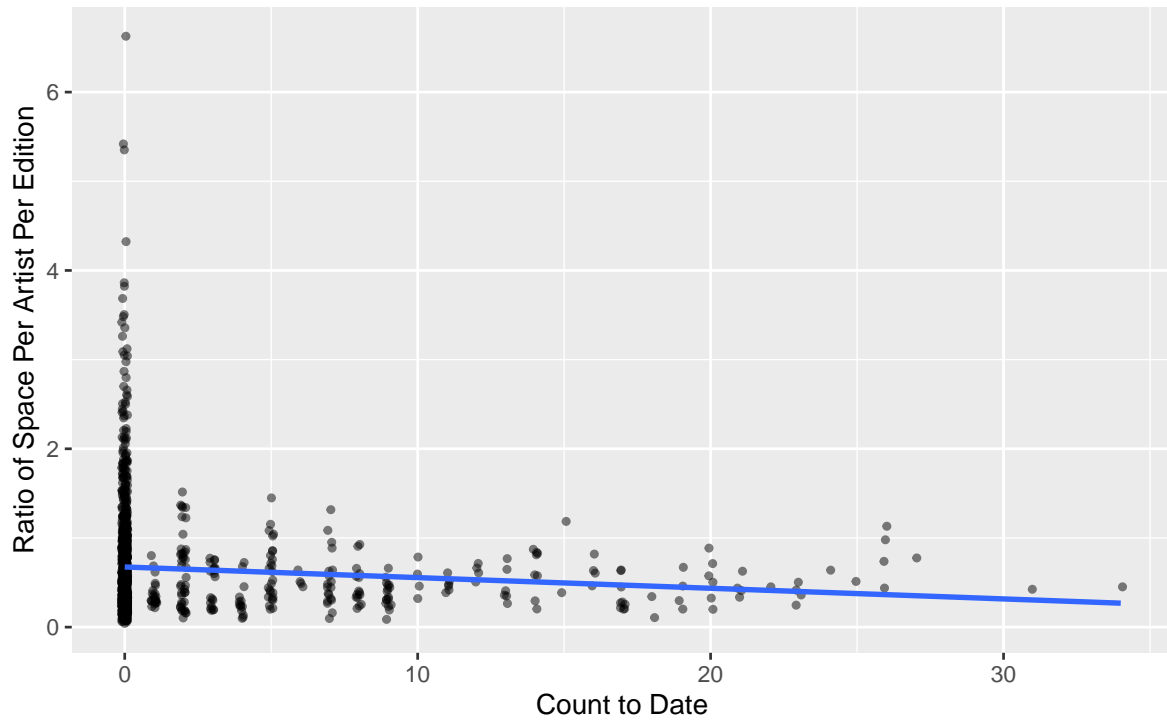
```
#corr_coef
```

```
ggplot(gardnerjanson_museums %>% filter(book == "janson"), aes(x = whitney_count_to_date, y = 
  geom_point(position=position_jitter(h=0.1, w=0.1), alpha = 0.5, size = 1) + 
  geom_smooth(method = "lm", se = FALSE)+ 
  labs(
    title = "Ratio of Space Per Artist Per Edition in Janson by\nNumber of Exhibitions per A",
    x = "Count to Date",
    y = "Ratio of Space Per Artist Per Edition"
  )
)
```

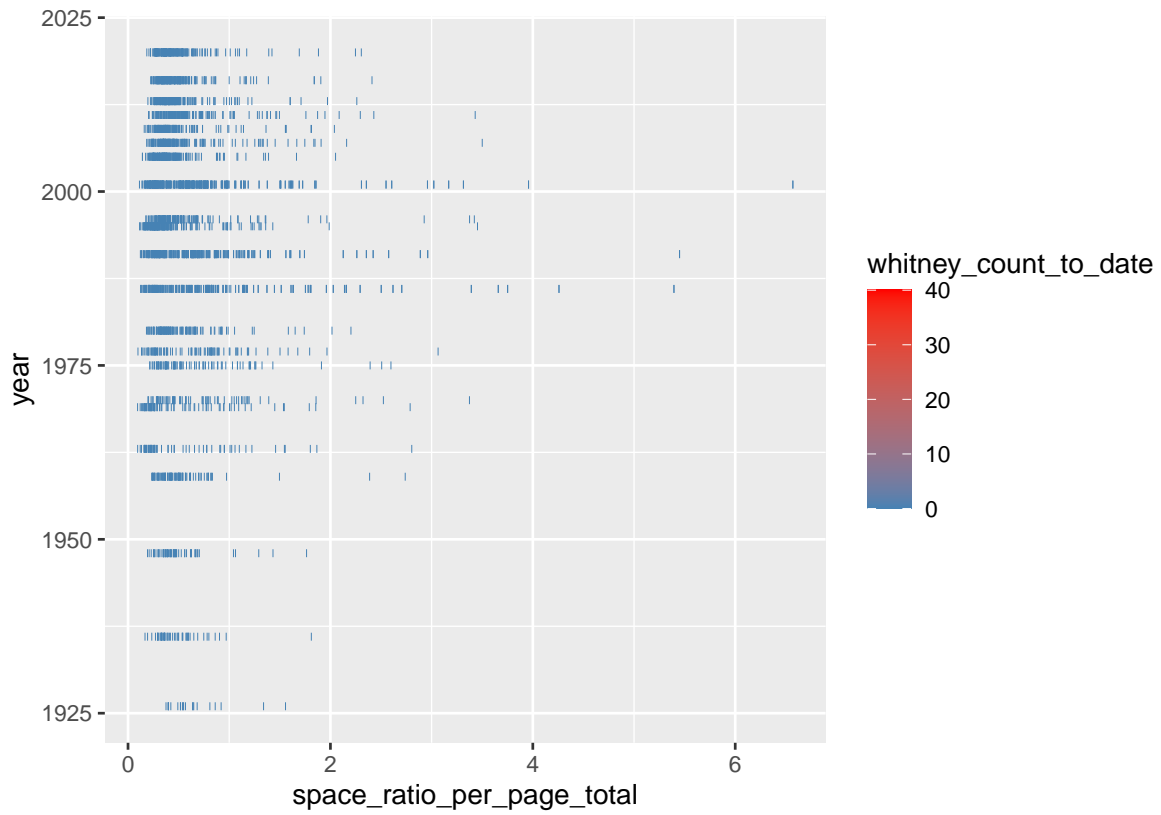
```
`geom_smooth()` using formula 'y ~ x'
```



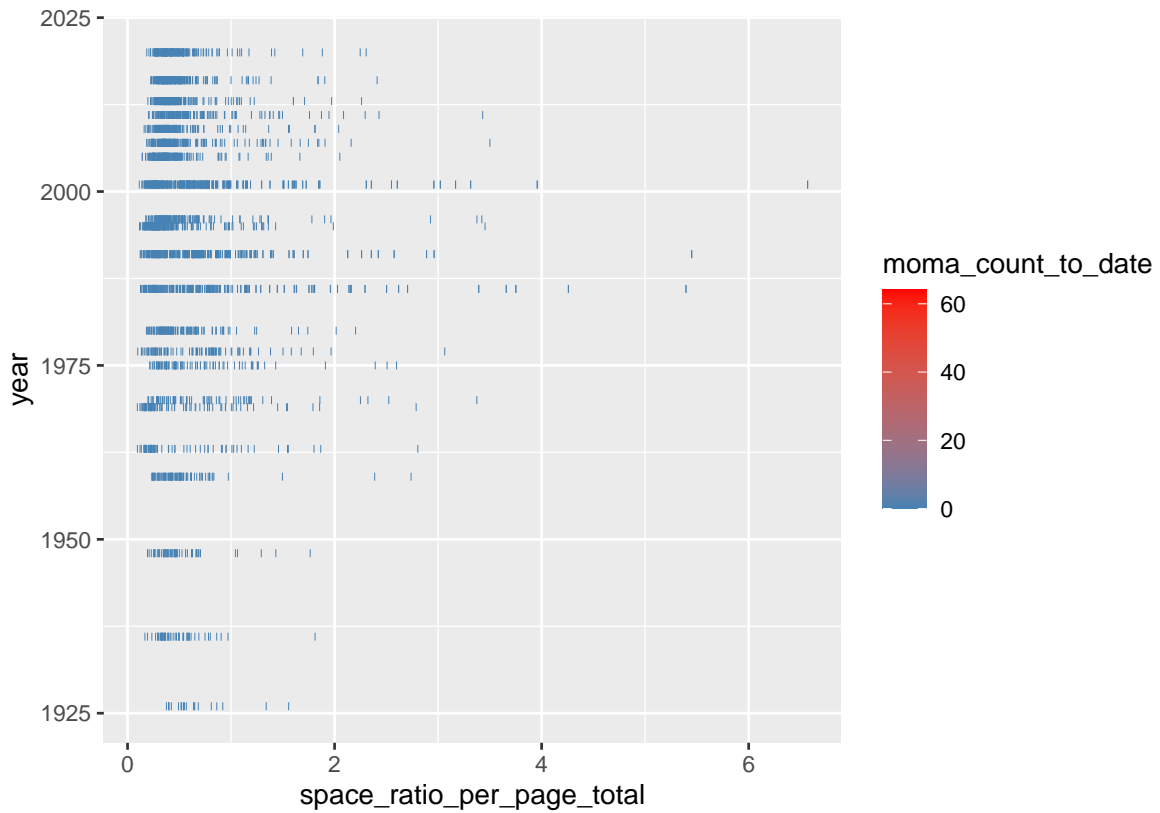
Ratio of Space Per Artist Per Edition in Janson by  
Number of Exhibitions per Artist per Editions in the Whitney



```
p <- ggplot(gardnerjanson_museums, aes(space_ratio_per_page_total,year)) +
  geom_tile(aes(fill = whitney_count_to_date), colour = "steelblue") +
  scale_fill_gradient(low = "steelblue",high = "red")
p
```

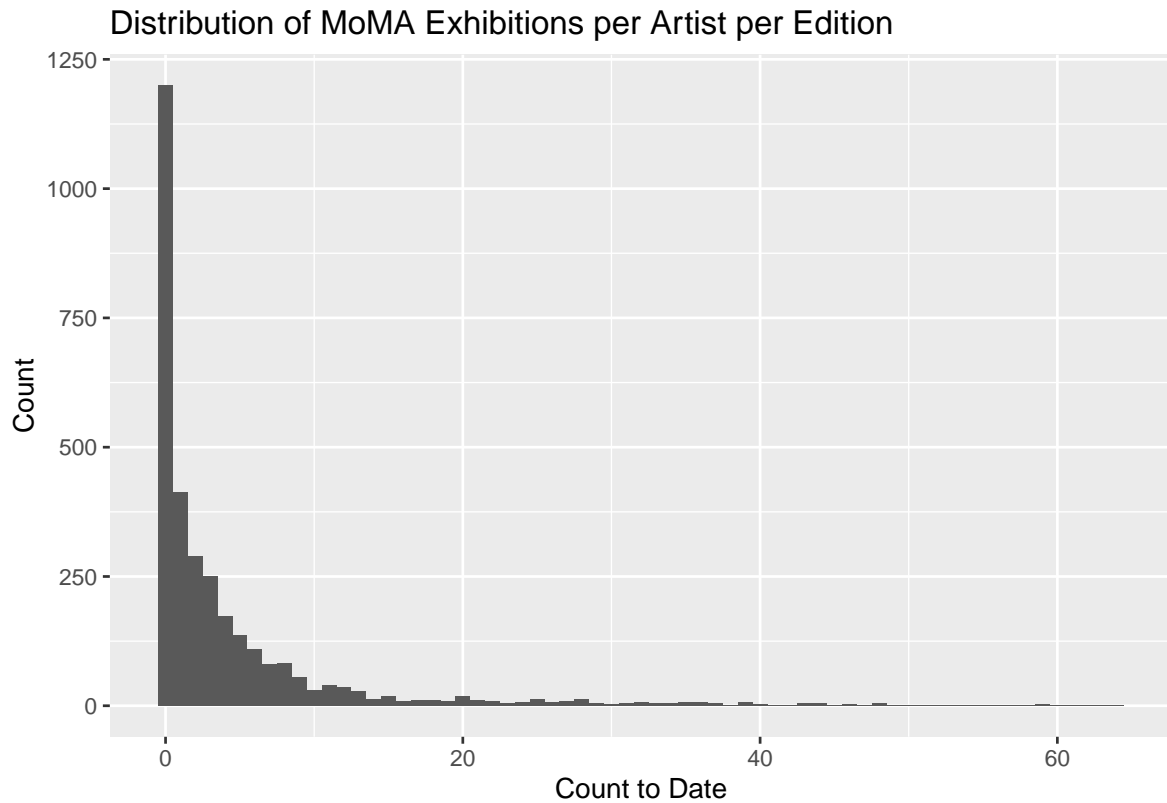


```
g <- ggplot(gardnerjanson_museums, aes(space_ratio_per_page_total, year)) +
  geom_tile(aes(fill = moma_count_to_date), colour = "steelblue") +
  scale_fill_gradient(low = "steelblue", high = "red")
g
```



moma\_eda

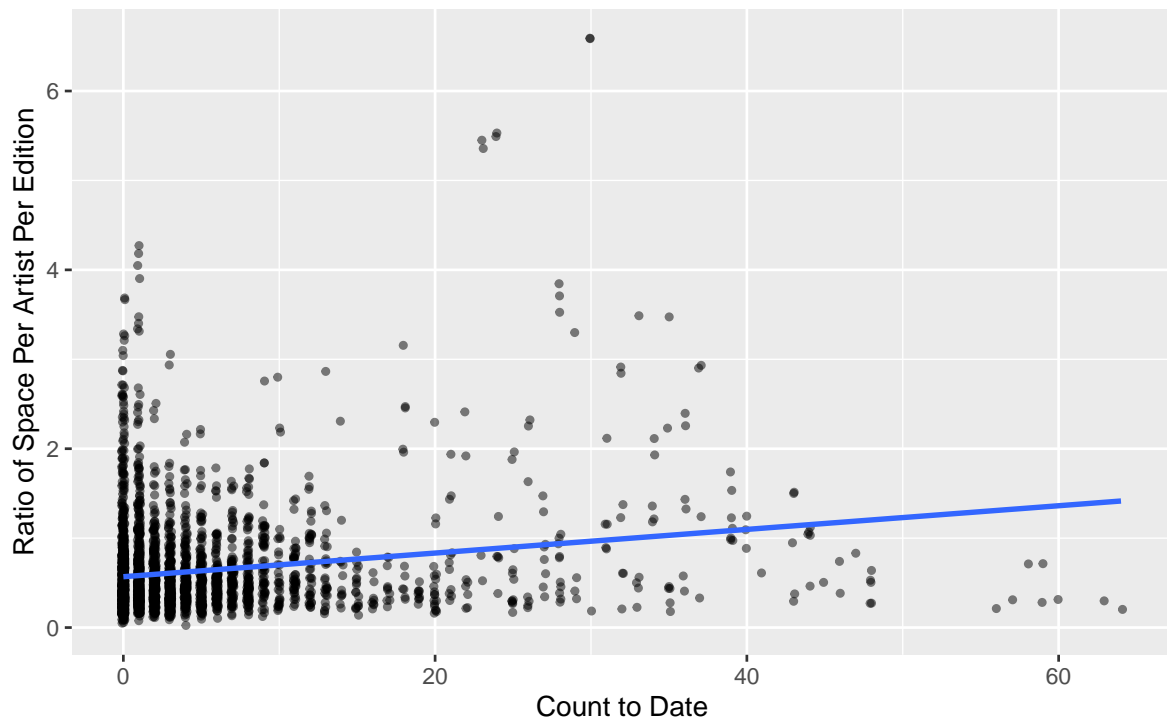
```
ggplot(gardnerjanson_museums, aes(x = moma_count_to_date)) +
  geom_histogram(binwidth = 1) +
  labs(
    title = "Distribution of MoMA Exhibitions per Artist per Edition",
    x = "Count to Date",
    y = "Count"
  )
```



```
ggplot(gardnerjanson_museums, aes(x = moma_count_to_date, y = space_ratio_per_page_total)) +
  geom_point(position=position_jitter(h=0.1, w=0.1), alpha = 0.5, size = 1) +
  geom_smooth(method = "lm", se = FALSE)+
  labs(
    title = "Ratio of Space Per Artist Per Edition by\nNumber of Exhibitions per Artist per l",
    x = "Count to Date",
    y = "Ratio of Space Per Artist Per Edition"
  )
```

`geom\_smooth()` using formula 'y ~ x'

# Ratio of Space Per Artist Per Edition by Number of Exhibitions per Artist per Editions in the MoMA



```
corr_coef_moma <- cor(gardnerjanson_museums$moma_count_to_date,
                      gardnerjanson_museums$space_ratio_per_page_total,
                      use = "complete.obs")
```

```
gardnerjansonSLR <- lm(space_ratio_per_page_total ~
  artist_race,
  data = gardnerjanson_museums
)
summary(gardnerjansonSLR)
```

Call:

```
lm(formula = space_ratio_per_page_total ~ artist_race, data = gardnerjanson_museums)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.5469	-0.2990	-0.1519	0.0952	5.9281

Coefficients:

	Estimate	Std. Error
(Intercept)	0.49076	0.15682
artist_raceAsian	-0.09472	0.16831
artist_raceBlack or African American	-0.06011	0.16778
artist_raceN/A	-0.15331	0.18647
artist_raceNative Hawaiian or Other Pacific Islander	-0.03479	0.19346
artist_raceWhite	0.15073	0.15714

	t value	Pr(> t )
(Intercept)	3.129	0.00177 **
artist_raceAsian	-0.563	0.57365
artist_raceBlack or African American	-0.358	0.72016
artist_raceN/A	-0.822	0.41105
artist_raceNative Hawaiian or Other Pacific Islander	-0.180	0.85730
artist_raceWhite	0.959	0.33755

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5433 on 3156 degrees of freedom  
Multiple R-squared: 0.01203, Adjusted R-squared: 0.01047  
F-statistic: 7.689 on 5 and 3156 DF, p-value: 3.393e-07

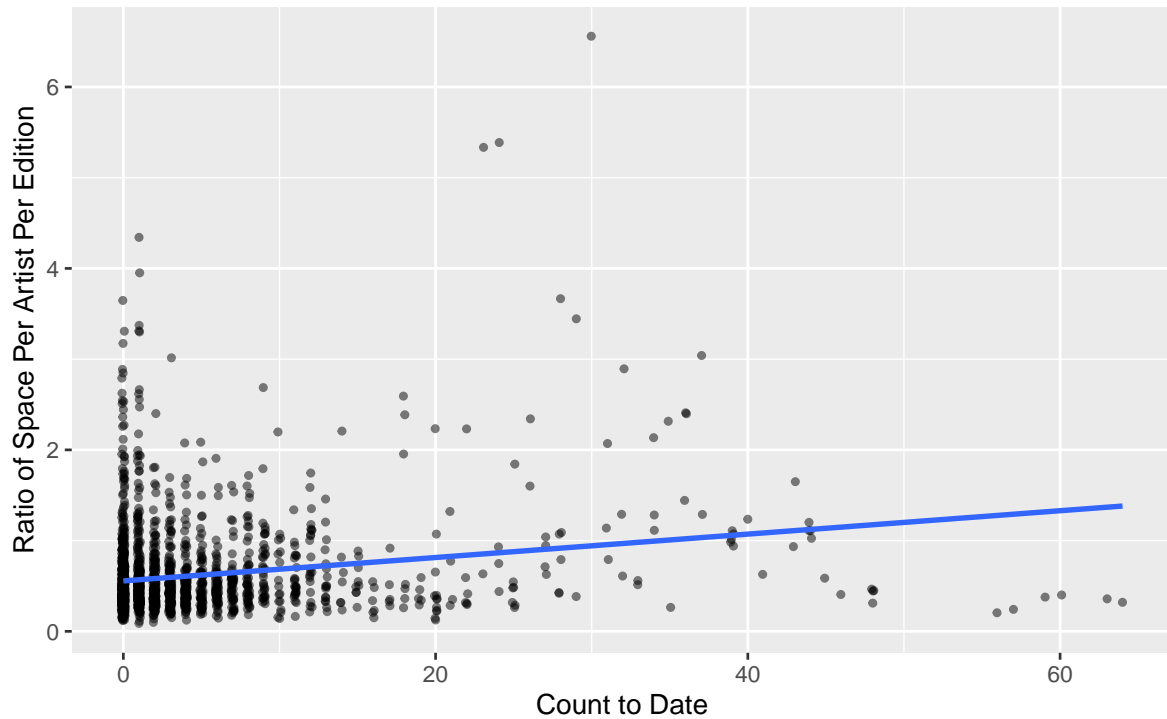
```
glance(gardnerjansonSLR)$r.squared
```

```
[1] 0.01203452
```

```
ggplot(gardnerjanson_museums %>% filter(book == "gardner"), aes(x = moma_count_to_date, y = ratio_space_per_artist_per_edition)) +
  geom_point(position=position_jitter(h=0.1, w=0.1), alpha = 0.5, size = 1) +
  geom_smooth(method = "lm", se = FALSE)+
  labs(
    title = "Ratio of Space Per Artist Per Edition in Gardner by\nNumber of Exhibitions per Artist",
    x = "Count to Date",
    y = "Ratio of Space Per Artist Per Edition"
  )
```

```
`geom_smooth()` using formula 'y ~ x'
```

Ratio of Space Per Artist Per Edition in Gardner by  
Number of Exhibitions per Artist per Editions in the MoMA



```
ggplot(gardnerjanson_museums %>% filter(book == "janson", moma_count_to_date != 0), aes(x = moma_count_to_date, y = ratio_of_space_per_artist_per_edition)) +
  geom_point(position=position_jitter(h=0.1, w=0.1), alpha = 0.5, size = 1) +
  geom_smooth(method = "lm", se = FALSE)+
  labs(
    title = "Ratio of Space Per Artist Per Edition in Janson by\nNumber of Exhibitions per Artist",
    x = "Count to Date",
    y = "Ratio of Space Per Artist Per Edition"
  )
```

`geom\_smooth()` using formula 'y ~ x'

Ratio of Space Per Artist Per Edition in Janson by  
Number of Exhibitions per Artist per Editions in the MoMA

