Lab 07 - UoE Art Collection

Tensorflow2.0

10/22/2020

Load packages and data

```
library(tidyverse)
library(skimr)
```

```
uoe_art <- read_csv("data/uoe-art.csv")
```

Exercise 10

At first, lets separate the year from our title column.

```
uoe_art <- uoe_art %>%
  separate(title, into = c("title", "date"), sep = " \\(") %>%
  mutate(year = str_remove(date, "\\)") %>%
  as.numeric()) %>%
  select(title, artist, year, link)
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 81 rows [27, 34, 35, 37, 38, 50, 96, 111,
```

```
## Warning: Problem with `mutate()` input `year`.
## i NAs introduced by coercion
## i Input `year` is `str_remove(date, "\\)") %>% as.numeric()`.
```

```
## Warning in function_list[[k]](value): NAs introduced by coercion
```

The first warning was generated because there were other pieces of information in the title of some of the art pieces which we could have stored in other columns but were ignored. We can ignore this warning because we are just concerned about the title and the year.

The second warning was generated because in some cases, we tried to convert characters into numeric value. We can ignore this warning as well because those characters were replaced with NA.

The third warning was probably generated to indicate that NA were being input in some function that we used above. We can ignore this as well because this will not affect our aim to separate title into two different columns.

Exercise 11

Now, lets print a summary of the dataframe.

```
skim_without_charts(uoe_art)
```

Table 1: Data summary

Name	uoe_art
Number of rows	2970
Number of columns	4
Column type frequency:	
character	3
numeric	1
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
title	0	1.00	2	95	0	1336	0
artist	115	0.96	2	55	0	1112	0
link	0	1.00	57	60	0	2970	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
year	765	0.74	1979.52	52.88	2	1958	1968	2018	2020

123 pieces are missing artist information while as 771 pieces are missing year information.

Exercise 12

Now, lets visualize the distribution of years of the pieces.

```r
uoe_art %>%
  ggplot(aes(x = year)) +
  geom_histogram() +
  labs(
    x = "Year",
    y = "Number of Pieces",
    title = "Distribution of Years of Art Pieces"
  )
```

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

Warning: Removed 765 rows containing non-finite values (stat_bin).

Distribution of Years of Art Pieces

The histogram has a value at around 0 AD which is far ahead than the year of all the other art pieces. In other words, the histogram indicates the presence of an outlier.

Exercise 13

The art piece titled "Death Mask (2)" had the out-of-the-ordinary information. Our code couldn't capture the correct information because it took the brackets in the title "(2)" itself as the date and ignored the actual date that came after the "(2)".
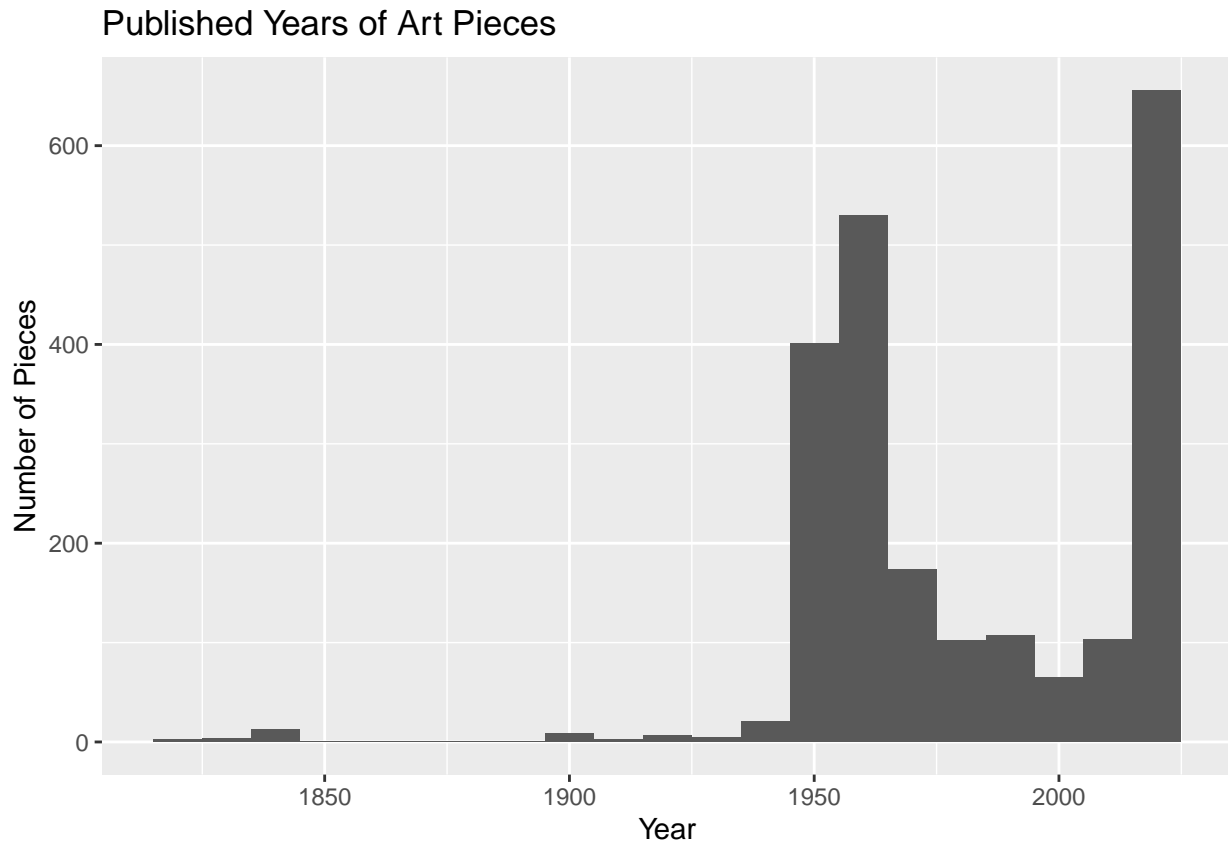
Lets correct the year information of the outlier that we mentioned above.

```
uoe_art <- uoe_art %>%
   mutate(year = case_when(
   year == 2  ~ 1964,
   TRUE       ~ year
 ))
```

Now, lets visualize the distribution of the years of the pieces again.

```
uoe_art %>%
  ggplot(aes(x = year)) +
  geom_histogram(binwidth = 10) +
  labs(
    x = "Year",
    y = "Number of Pieces",
    title = "Published Years of Art Pieces"
  )
```

```
## Warning: Removed 765 rows containing non-finite values (stat_bin).
```

Published Years of Art Pieces

The distribution is bimodal with peaks at around 1960 and 2020 AD. While majority of art pieces had years clustered in between 1950 and 2020 AD, there were some art pieces that had year as early as around 1800 AD as well.

Exercise 14

Now, lets find the most commonly featured artist in the collection.

```
uoe_art %>%
  count(artist, sort = TRUE) %>%
  filter(artist != "Unknown" & artist != "NA")
```

```
## # A tibble: 1,111 x 2
##    artist              n
##    <chr>           <int>
##  1 Emma Gillies      148
##  2 John Bellany       22
##  3 Ann F Ward         19
##  4 Boris Bućan        17
##  5 Marjorie Wallace   17
##  6 Zygmunt Bukowski   17
##  7 Gordon Bryce       16
##  8 William Gillon     16
##  9 Robert Rivers      14
## 10 Denness Morton     13
## # ... with 1,101 more rows
```

The most commonly featured artist was Emma Gillies. She had 148 arts featured in the collection. She was

followed by John Bellany and Ann F Ward each with 22 and 19 arts featured respectively.

They must have been associated with the university in some way or the other because of which the university has so many pieces from them. Probably, they were either students or faculty members at the University of Edinburgh.

Exercise 15

Finally, lets find the number of arts with the word "child" in their titles. The information of the art pieces with "child" in their title are as follows:

```
title_child <- uoe_art %>%
    filter(
      str_detect(.$title, "Child") |
      str_detect(.$title, "child")
    )
```

```
knitr::kable(title_child)
```

title	artist	year	link
Untitled - Children Playing	Monika L I Uecker	1963	https://collections.ed.ac.uk/ art/record/99445?highlight=:
Child's collar. Chinese	Unknown	2018	https: //collections.ed.ac.uk/art/ record/102674?highlight=:
The Sun Dissolves while Man Looks Away from the Unborn Child Mesmerized by the New Technologies	Eduardo Luigi Paolozzi KBE, RA	2018	https://collections.ed.ac.uk/ art/record/21703?highlight=:
Virgin and Child	Unknown	NA	https://collections.ed.ac.uk/ art/record/20544?highlight=:
Virgin and Child	Unknown	2018	https://collections.ed.ac.uk/ art/record/20571?highlight=:
Virgin and Child	Unknown	NA	https://collections.ed.ac.uk/ art/record/20559?highlight=:
Untitled - Portrait of a Woman and Child	William Gillon	NA	https://collections.ed.ac.uk/ art/record/99481?highlight=:
Child's chinese headdress	Unknown	2018	https: //collections.ed.ac.uk/art/ record/102677?highlight=:
Figure Composition with Nurse and Child, and Woman with Budgies	Edward A. Gage	NA	https://collections.ed.ac.uk/ art/record/22579?highlight=:
Woman with Child and Still Life	Catherine I. McIvor	1938	https://collections.ed.ac.uk/ art/record/22512?highlight=:
The Children's Hour	Eduardo Luigi Paolozzi KBE, RA	NA	https://collections.ed.ac.uk/ art/record/21699?highlight=:

There are 11 art pieces that have a word containing the string "child" in their title.