

# Homework 1 Priceless Art

AUTHOR

Marta Grinberga

## Where is this data from?

---

**Who** : 7 famous artists and 7399 less known artists.

**What** : Art being listed at an auction.

**Why** : The data is a part of an artist's science project, that is meant to show how the commercial art auction market evaluates art pieces.

**Where** : Artists are from all over the world

**When** : 2014 is the last entry for listings and backward.

**How** : Data is gathered from an auction website and cleaned.

## What are the variable types?

---

artist (nominal)

country (categorical)

yearOfBirth (quantitative, years)

name (nominal)

year (quantitative, years)

ageOfPainting (quantitative, years)

price (quantitative, \$)

material (categorical)

height (quantitative, inches)

dominantColor (categorical)

## Filtering for art from China

```
chinaart<- artdata%>%  
  filter(country=="Chinese")
```

## The distribution of the height of the Chinese paintings

I will investigate the distribution of the height of 874 paintings created by Chinese artists. The oldest painting dates back to 11th century (out of the paintings with known year of painting), and the most recent to 2014.

Chinese art is usually associated with minimalistic techniques of expression, but does this mean that the art should also convey a small form?

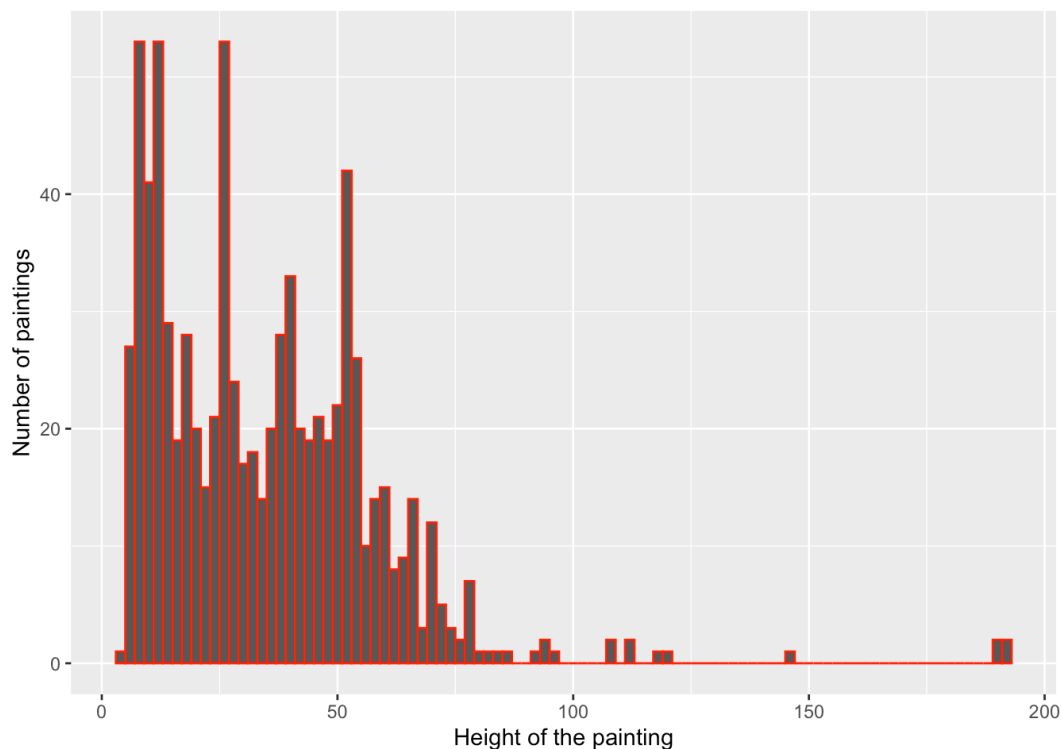


Figure 1: The distribution of the height of the Chinese paintings in a histogram

The histogram displays the needed information about the height of the paintings in an organized way, and displays 3 modes of approximately 8, 12 and 26 inches. This histogram is left-skewed and contains most of its elements within the 75 inch range. The mean for this data is 34.7 inches.

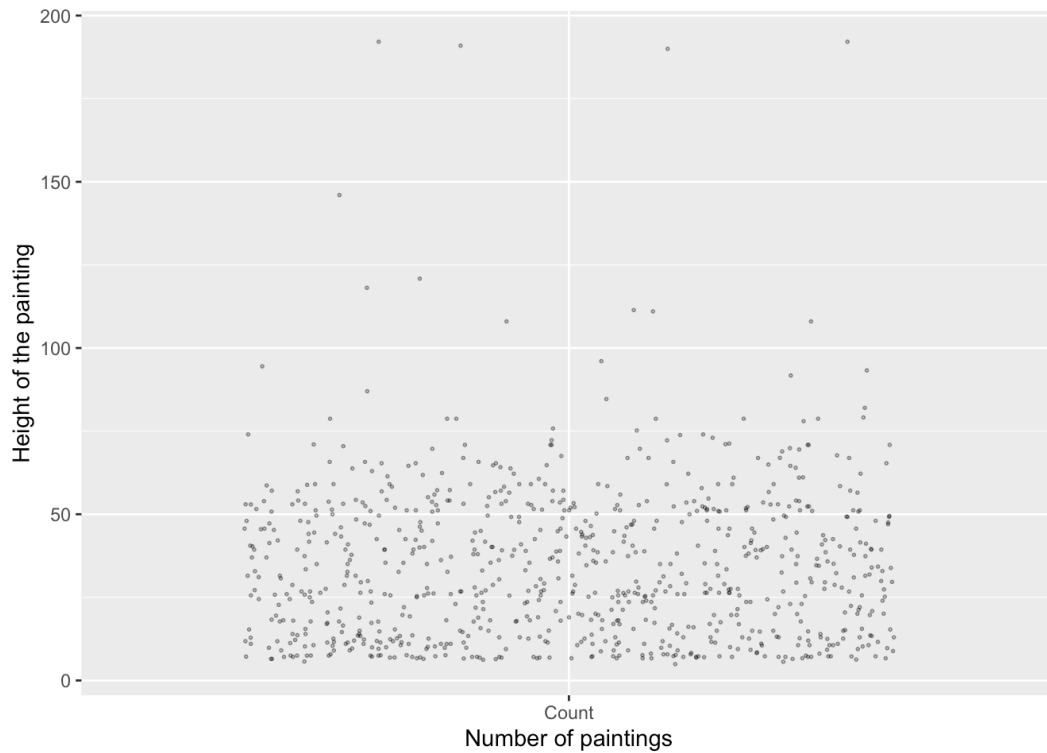
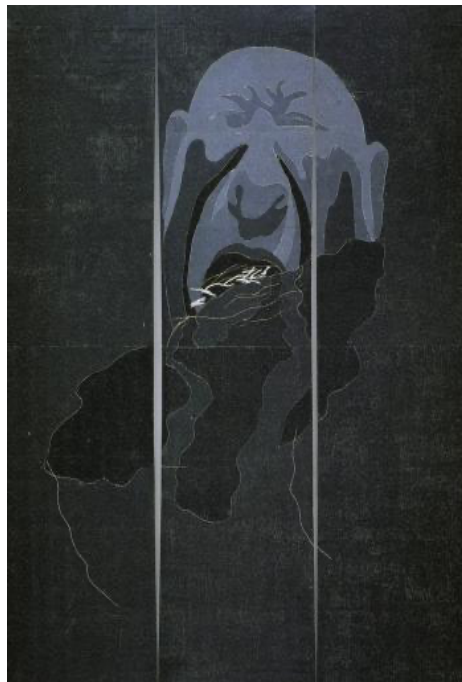


Figure 2: The distribution of the height of the Chinese paintings in a scatter plot

The scatter plot, similarly to histogram shows that the height of paintings is densely contained within 75 inches. There are some far outliers:

3 of the furthest outliers are paintings by artist Fang Lijun, whose height is around 191-192 inches:





These works are heavily inspired by social and political events. They are bold and one might call them "brutal". The price for them range from 50'000 to 141'000\$.

## Conclusion

Most of the paintings in this data set have a relatively small height with some far outliers that prove that Chinese art is diverse.

## The distribution of the width of the Chinese paintings

---

Height of the painting is only a half of its dimensional properties. To get the full image, the same analysis will be done with width.

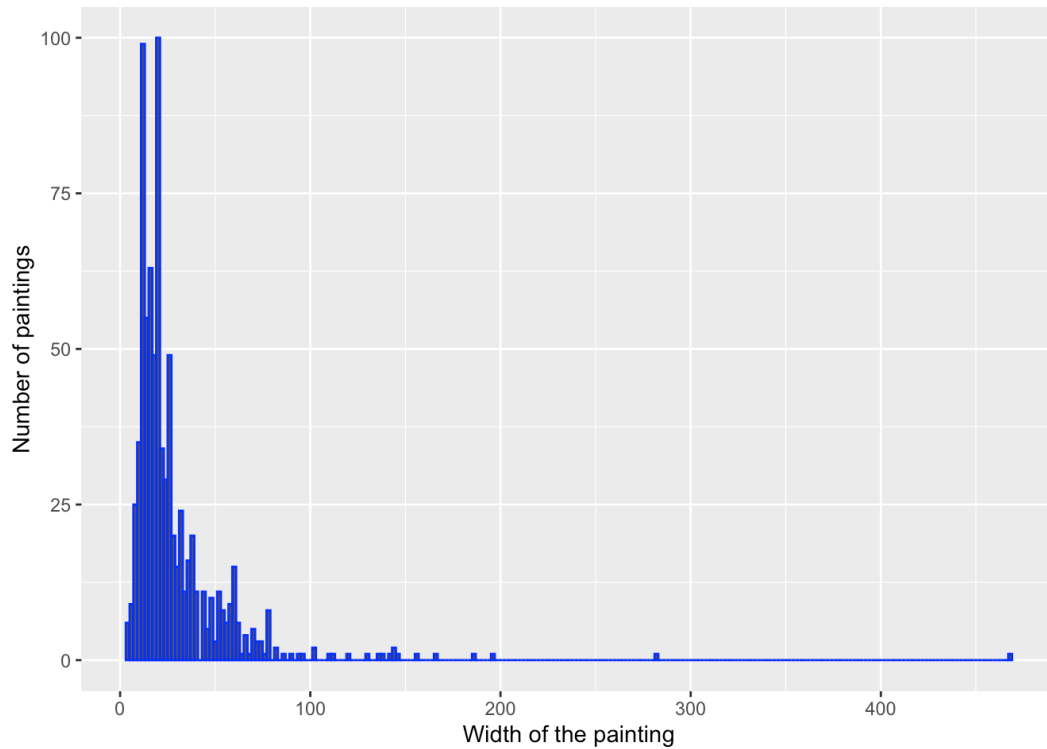


Figure 3: The distribution of the width of the Chinese paintings in a histogram

The mean for the width of Chinese paintings is: 28.6 inches.

Similarly to the height of paintings, the histogram of the width is left-skewed, with two modes. There is one "extreme" outlier, which will be investigated...

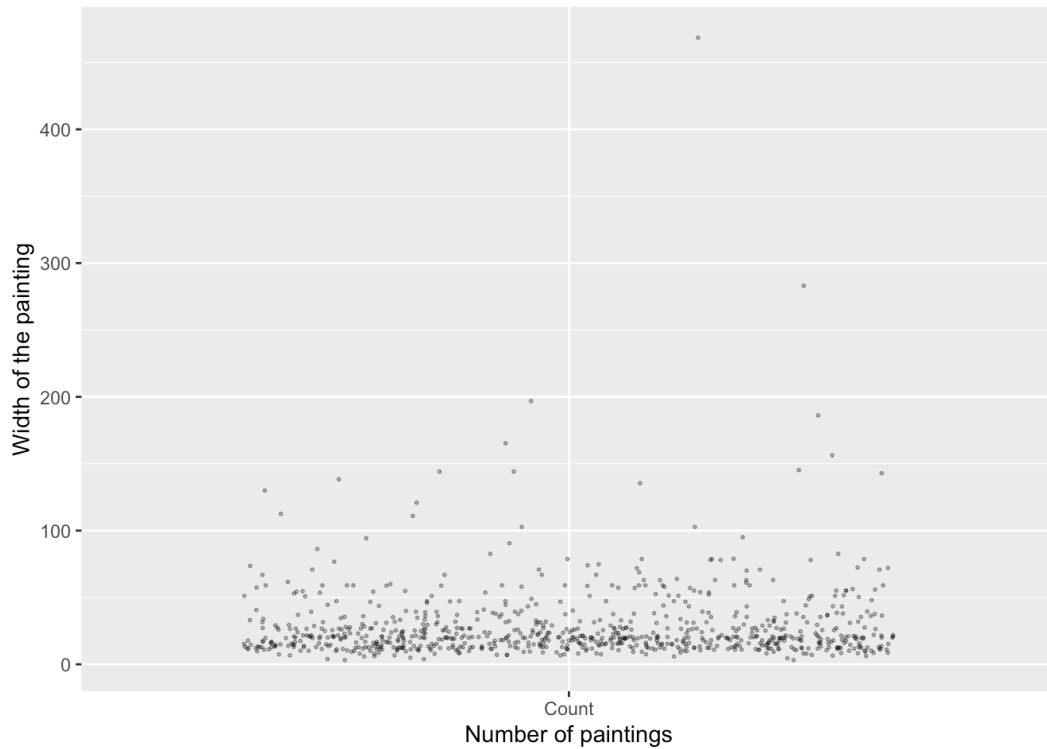


Figure 4: The distribution of the width of the Chinese paintings in a scatter plot

The scatter plot shows that most of the paintings are mostly within the 50 inch range and the furthest outlier lays beyond the mark of 450 inches.

### Painting "Grand View of Mountains and Rivers" by Dai Jin (1388-1462)



This landscape painting came to life during the Ming dynasty. This style of painting dates back to 6th century and is an important part of Chinese art and culture. Having this painting in the data set indicates about its diversity and inclusiveness.

## Conclusion

The width of the paintings is mostly contained within 50 inches, which is a little smaller than an A1 paper size.

## Height and width combined

---

Most of the paintings created by Chinese artists can be described as "small", due to a historic influence of Buddhism and Persian culture. The art of miniature paintings has continued for centuries and has many followers in the 21st century, which is shown by the left-skewed distributions of the graphs. The outliers investigated point out to the diversity of the data set - paintings of landscape from 13th century and modern art that represents the social and political problems from 20th century. I can also point out that the mean for the width of the paintings is by around 6 inches larger than that of the height, which might indicate that paintings have a more horizontal distribution, or it might be due to several horizontal paintings in the data set.

## Relationships between categorical variables - American and Chinese artists and oil vs. ink

---

### Recording my data

```
library(dplyr)
ELSE <- TRUE
artdatamutated <- artdata %>%
  mutate(
    material.type = case_when(
      grepl("Oil", material, ignore.case = TRUE) ~ "oil-ish",
      grepl("Ink", material, ignore.case = TRUE) ~ "ink-ish",
      is.na(material) ~ "Unknown",
      TRUE ~ "Other"
    ),
    us.china = case_when(
      grepl("USA|United States|America", country, ignore.case = TRUE) ~ "USA",
      grepl("China|Chinese", country, ignore.case = TRUE) ~ "China",
      TRUE ~ "Other"
```

```
)  
)
```

Warning: There were 4 warnings in `mutate()`.

The first warning was:

i In argument: `material.type = case\_when(...)`.

Caused by warning in `grepl()`:

! unable to translate 'etching\_on\_Arches\_<c3>' to a wide string

i Run `dplyr::last\_dplyr\_warnings()` to see the 3 remaining warnings.

## Investigating the categorical relationship between **us.china** and **material.type**

I decided to first create a table that will display all of the data, making the numbers more accessible and understandable.

	ink-ish	oil-ish	Other
American	1820	1993	6876
Chinese	585	152	138
Other	1027	11063	16416

The first noticeable difference between American and Chinese artists is the materials they use - Chinese artists (in this data set) prefer ink to oil and other materials, it being 67% in this investigation. Americans, on the other hand use a variety of other materials over ink and oil, which have a similar amount.

To make a clearer image, I used a bar chart to display the relationship in visually.



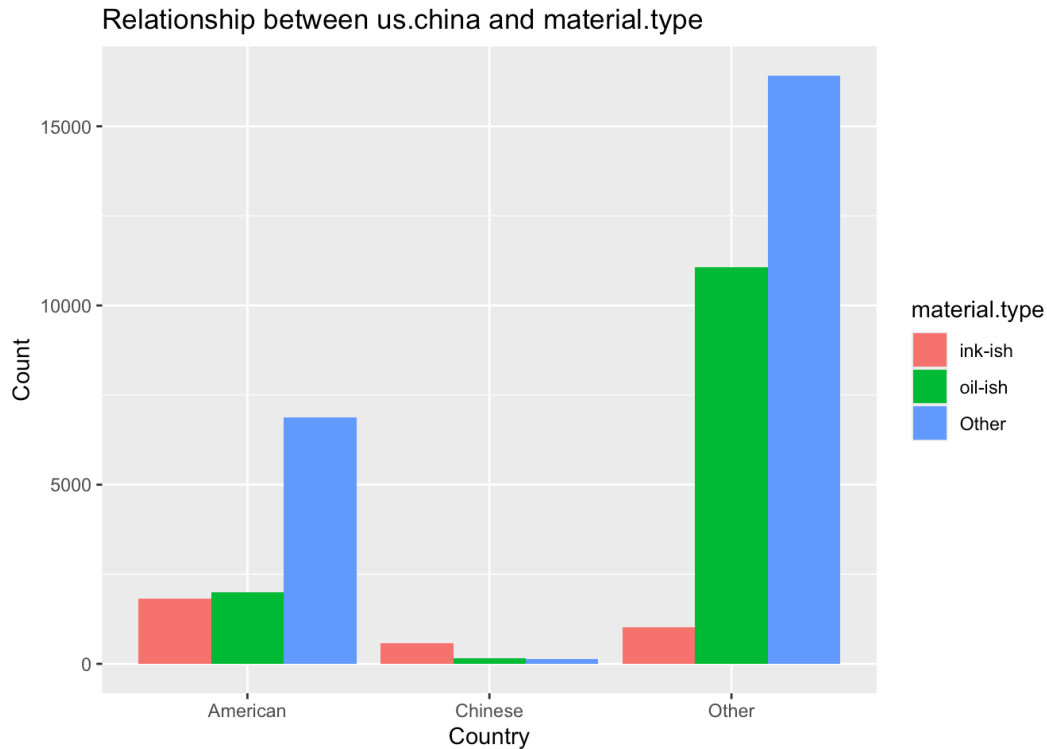


Figure 5: The categorical relationship between us.china and material.type

From this chart we can see that artists from other countries use oil and other materials much more than Americans and Chinese.

## Conclusion

After comparing the materials, my predictions about Chinese artists using ink in their paintings came true, because it is most commonly used for landscape paintings, which are one of the most known forms of art in China. Historic fact: The ink paintings first appeared during the Tang dynasty (7th-9th century) and developed in China ever since. As for the American artists, they do not use "traditional" materials as much. One can refer to the movements of pop art or abstract expressionism. Artists from other countries can hardly be called a category, because they vary in cultural background, religion and beliefs, which can be factors for different use of materials (charcoal, watercolors, gouache...).

## Comparing, comparing, comparing

### Recoding my data

For this exercise I will use the data of price and area (that can be

calculated as the height of the painting \* the width of the painting) for 4 of the "popular" countries(The US, France, Italy and Spain).

```
artdatamutatedtwo <- artdata %>%
  mutate(famous.countries = case_when(
    grepl("America", country, ignore.case = TRUE) ~ "American",
    grepl("French", country, ignore.case = TRUE) ~ "French",
    grepl("Italy|Italian", country, ignore.case = TRUE) ~ "Italian",
    grepl("Spanish", country, ignore.case = TRUE) ~ "Spanish",
    TRUE ~ NA_character_ ), area = height * width)
```

## Comparison of the prices of paintings

To see the different distribution of prices per country, I used a scattered plot:

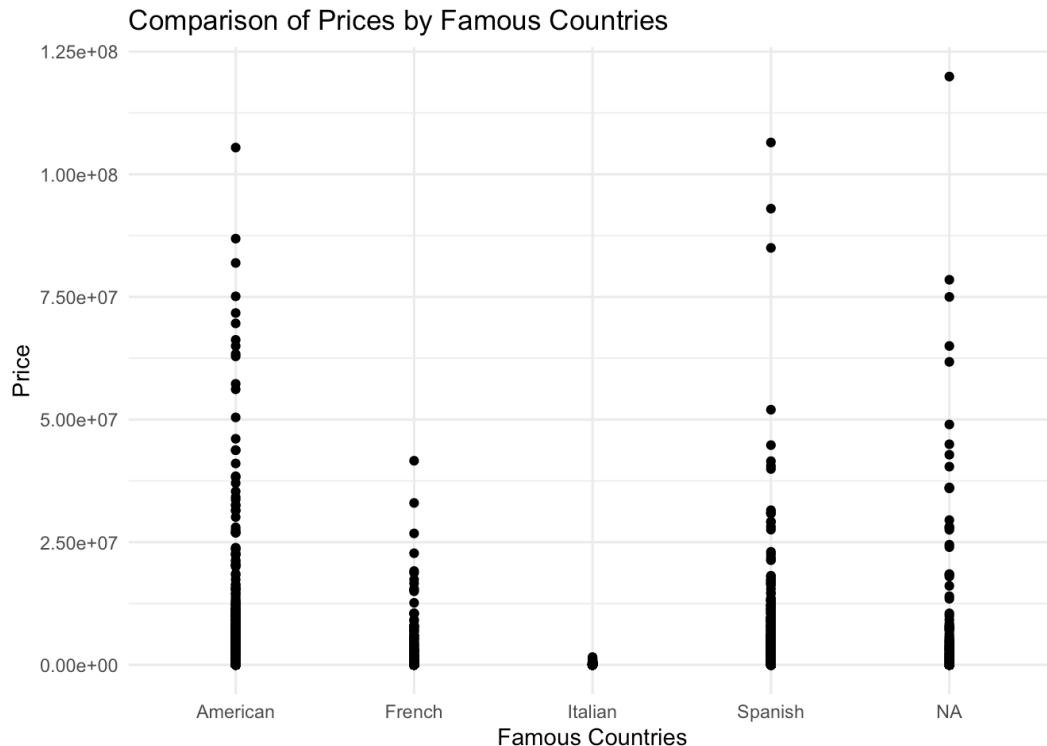


Figure 6: The comparison of prices per painting in the popular countries

At the first glance, it seemed that there was a mistake made for the data of Italian artists, but after the investigation, it turned out that the most expensive Italian painting in the data base cost "only" 1'583'023\$, which is much lower than many other paintings in the data base. In my opinion, this drastic difference is not due to the quality of art, but acquisition of it by the art sellers. The auctioneers possibly were not able to get a hold or

permission of some of the finest Italian paintings.

As can be seen in the graph, the furthest outlier belongs to an artist, who does not come from the "famous" countries. Most of the artworks are contained within the price of 250K\$, lowest reaching only 3\$, which is a color silkscreen.

**The most expensive American painting in the data set is:**

Silver Car Crash (Double Disaster) by Andy Warhol (105'445'000\$)



**The most expensive Spanish painting in the data set is:**

Nude, Green Leaves and Bust by Pablo Picasso (106'482'500\$)



**The most expensive French painting in the data set is:**

Les Pommes by Paul Cezanne (41'605'000\$)



**The most expensive Italian art piece in the data set is:**

Piede by Luciano Fabro (polished bronze and silk)



**The most expensive art piece in the data set is:**

The Scream by Norwegian artist Edvard Munch (119'922'500\$)



## Conclusion

This data set contains very valuable and expensive artworks by world-renown artists. The artworks drastically differ in price and have the amplitude of 119'922'497\$. One could find art at any price, but should be careful with the type of art they get, because the cheapest artworks are often some type of print of known artists.

## Comparison of the areas of paintings

Similarly to the comparison of prices, I will create a scatter plot displaying different sizes of paintings. This will help determine if there is a relationship between price and area.

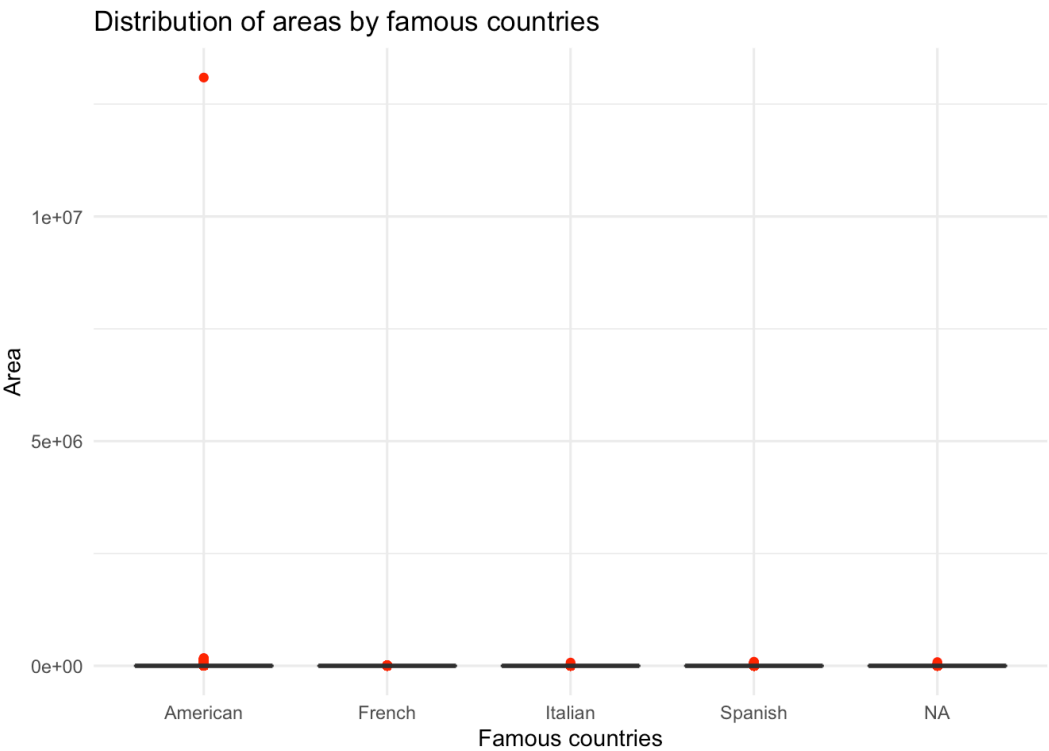


Figure 7: The comparison of areas of paintings in the popular countries

As can be seen, because of the far outlier (screen print of the painting “Franz Kafka” by Andy Warhol”) the data is not perceivable, but because it cannot be qualified as a mistake I will rescale the graph using logarithmic function and creating a box plot.



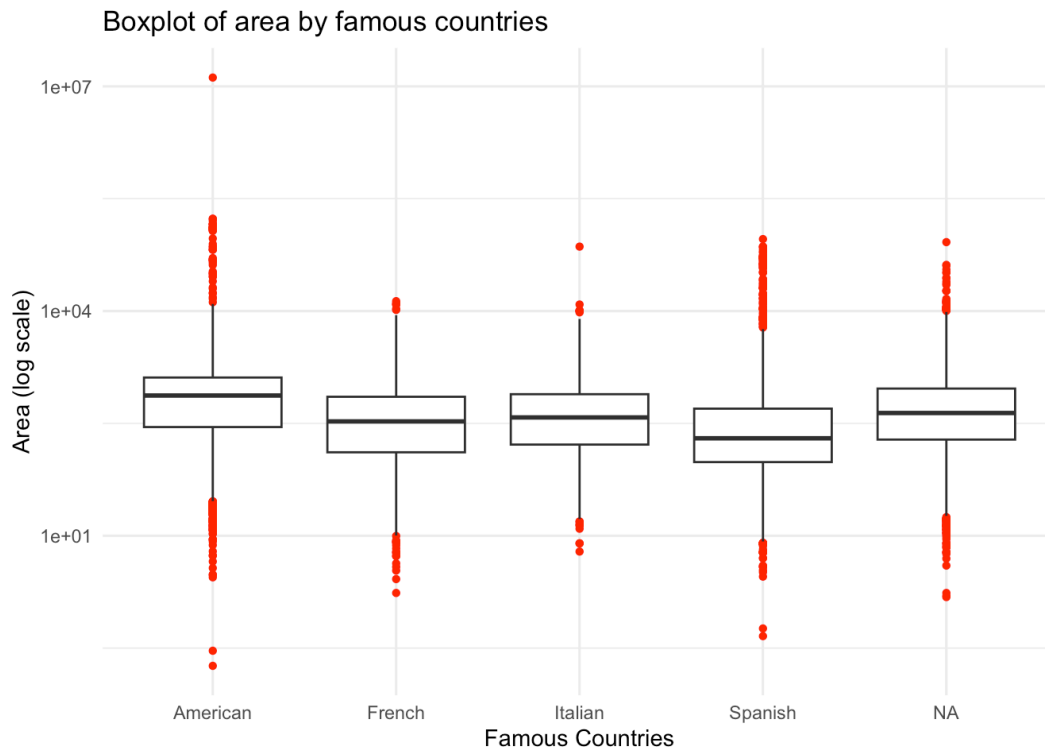


Figure 8: Boxplot of area by famous countries with logarithmic scale

Now it is easier to also identify and compare the data for each country. We can see that most of the paintings have a similar mean area and the IQRs are similarly sized.

The furthest lowest outliers are of American artworks, but after investigating the data set, I am fairly certain that the mistake was made in the recordings of height and width, because the furthest outlier (a painting by American artist Justus Dalee) has the area of 0,1833 inches<sup>2</sup>, which physically could not be possible, and after some more reaserching, I found out that the original size of the painting is 3 x 2.6 in. .

In accordance with acquired information, I would remove the lower outliers that cannot exist in a real world, where laws of physics hold.

## Differences and similarities of the art around the world

Although cultural differences between different nations might be big, the world of art allows a person, without prejudice create, paint and form. As we can see in the boxplot of painting areas, all of the countries have a very similar mean, which means that there might be a universal standard or understanding of proportion that comes into play, when artists create their



masterpieces. The biggest difference in the price distribution is the "low" prices for Italian art, which can be explained by incompleteness of the data set. In my opinion, these comparisons show that art does not have a preference over country and size. One is able to show their abilities to their maximum capacity by creating unique art and showcasing it to broad audience. The known paintings are appreciated by citizens of different countries, displayed in museums and auctioned worldwide.

## Brightness of paintings over time

---

To compare the brightness of artworks in the data set, I will choose 3 art movements from different periods in history :

Year 1766 (**Neoclassicism**)

Year 1885 (**Impressionism**)

Year 1942 (**Abstract Expressionism**)

I chose these art movements due to the limitations of the data set, and I tried to choose the most diverse data. As for the years 1766 and 1885 I chose them, because they were in the range of the art movements early stages, when the styles have not began to transform into new ones. And year 1942, because I am personally interested how WW2 impacted the paintings of artists - are they more dense, are they darker?

## Finding the average

```
art1766<- artdata%>%  
  filter(year=="1766")  
art1885<- artdata%>%  
  filter(year=="1885")  
art1942<- artdata%>%  
  filter(year=="1942")  
mean1766<-mean(art1766$brightness)  
mean1885<-mean(art1885$brightness)  
mean1942<-mean(art1942$brightness)  
mean1766
```

```
[1] 93.66667
```

```
mean1885
```

```
[1] 134.4746
```

```
mean1942
```

```
[1] 155.303
```

```
meanall<-mean(c(mean1766,mean1885,mean1942))  
meanall
```

```
[1] 127.8148
```

The brightness of paintings has increased consequently with each time period.

## Normalizing the data

	Year	Average_Brightness	Z_Score
1	1766	93.66667	-0.26716861
2	1885	134.47458	0.05210524
3	1942	155.30303	0.21506337

## Conclusion

The Z-score table helps to see that the standard deviation for paintings in each period is within 0.3 z-scores, so although it has increased overtime, we can see a consistent growth that can be attributed to different factors. The explanation can be found by addressing the definition of brightness of the data set, if it describes the brightness in the year data was collected, and not when the art piece is created, the diminishing brightness can be explained due to the aging of canvas and materials used. A painting of centuries ago, undergoes many physical dispositions. Some of the paintings can be renovated, but the original is bound to lose its color overtime.

## Relationship between the price of the painting and the creation year

It is commonly believed that paintings gain value overtime. But how true is

this? The idea for this inquiry comes from my friends studying art degrees and fearfully looking ahead. My goal is to determine if time or talent is the determining factor for the value of the painting.

First I will see the distribution of average prices of paintings throughout time.

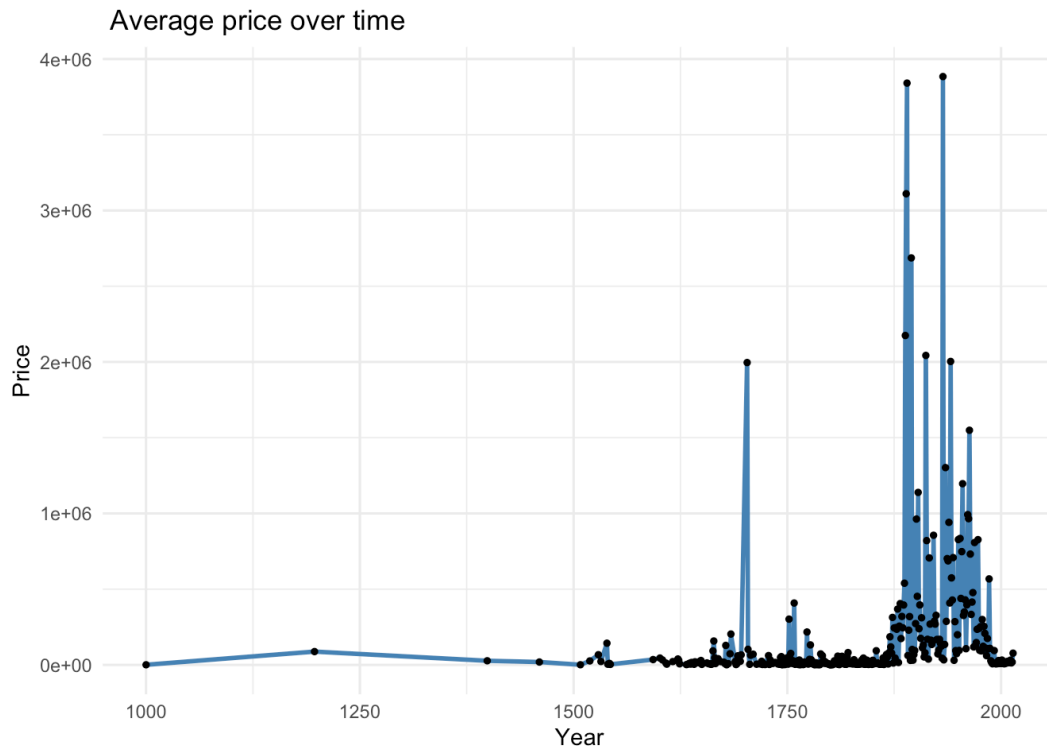


Figure 9: Distribution of prices throughout the year

As we can see from the graph, the most expensive paintings are all within the bounds of the last century, with another peak in 18th century (1703 painting by a Chinese artists Bada Shanren is listed at 3627480\$). These values are minimally affected by the unknown years of publications of older artworks, but I can conclude that the main shape would not change by much if they were included.

Secondly, I will investigate the prices of artworks in the last century (1900-2014).

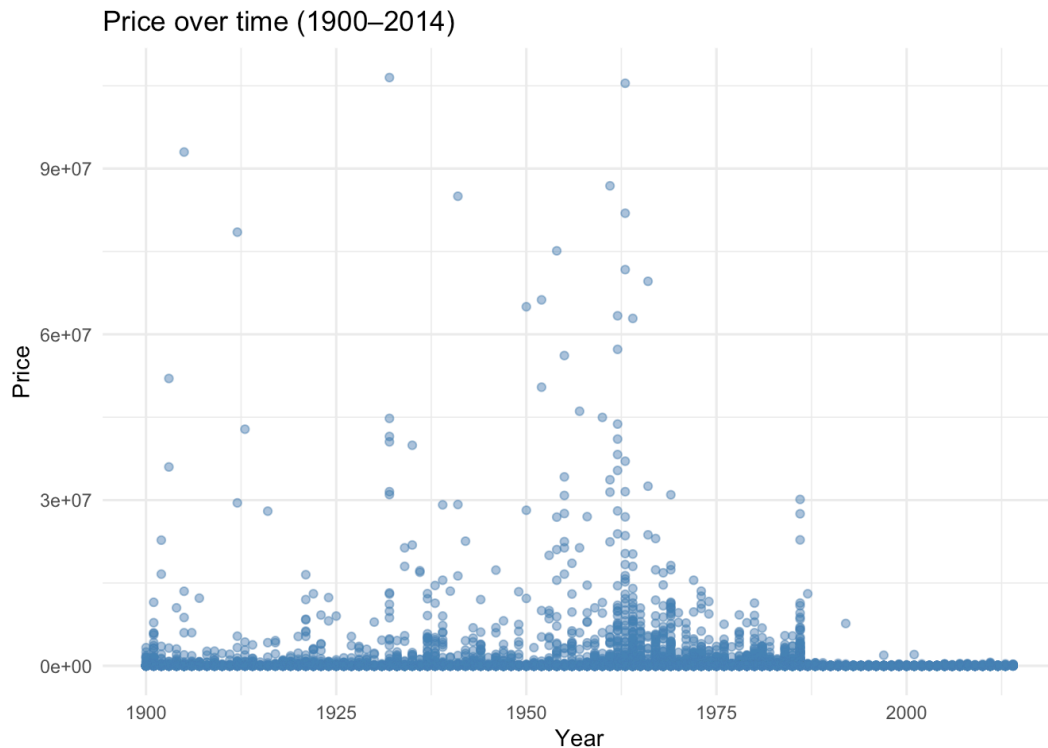


Figure 10: Distribution of prices throughout the years 1900–2014

This scatter plot clearly shows that prices for paintings are much higher for those pieces produced before 1990. This can be due to artists not having enough time to establish their names in the industry and showcase their art. Other explanation could be the rise of the technological advancements in 21st century. They most likely influence the way art looks, shape the personalities and realities of the creators and somewhat limit their potential.

At this point in my investigation, I cannot return to my friends with great news, but lastly I can try to soften the news by seeing what were the prices for artpieces listed from 1995 to the end of dataset. Maybe they will not become billionaires, but at least starving will not be a circulating factor.

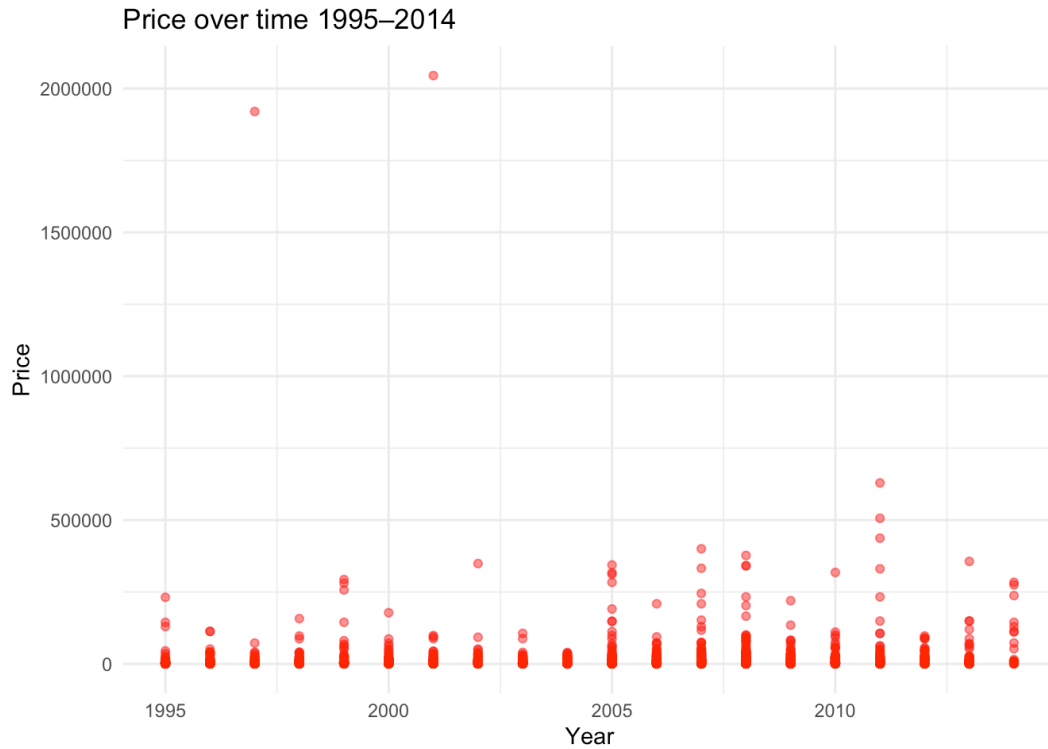


Figure 11: Distribution of prices throughout the years 1995–2014

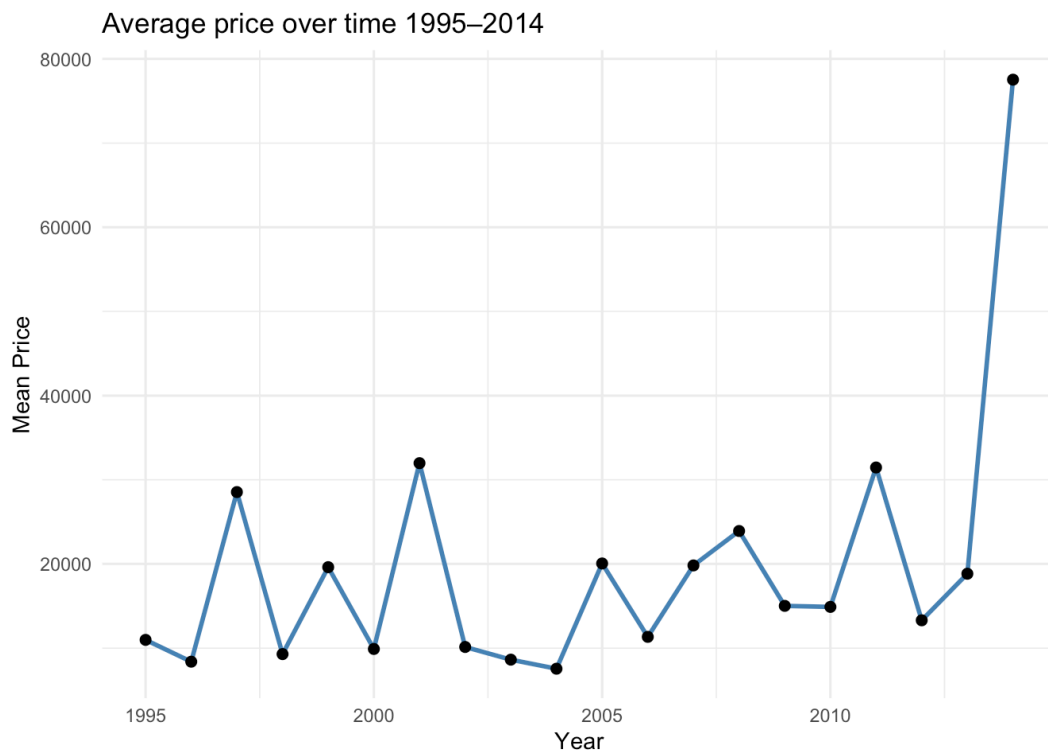


Figure 12: Distribution of prices throughout the years 1995–2014

These graphs are already showing more promise to new generation of artists. There is a far outlier in the year 2001, that turns out to be a bronze

sculpture by French artist Claude Lallemand (2001), sold for more than 2M\$



Another interesting outlier is an oil painting by Chinese artist Ma Baozhong, which is being sold 1'920'000\$. It's name is 19 December 1984, which is the date when China and the UK signed an agreement to transfer Hong Kong to China. It's area is more than 40'000 sq. inches.



Despite the fact that art auction prices peaked for art created in 1960s-1980s, I can conclude that the diversity of art in the data set is huge. For example, the photography collections are being sold, digital prints, prints, figurines, and other, as we can see in the distribution of materials. These artworks can be evaluated differently and are independent from paintings and vice versa.

Through this investigation, I can point out that ancient art is not evaluated better than that created in the last decade. I can also conclude that authenticity is one of the driving factors for success, because in the age of digital advancements soul is the key. And this advice I would also give to my artist friends, who are unsure of their future.

## Conclusion

---

Since the beginning of time art has been created around the world. Religious frescos, paintings in caves, miniature paintings , romantic interpretations of lide, and countless other forms of expression. Art is not what drives countries apart, it is what unites them to appreciate something other than materialistic thriving for wealth. Art takes all shapes and forms, and proves that realms of human brains are limitless. This investigation showed that artistic expression cannot be categorized and put a label on, it requires patience and devotion and gives back joy.