Artist Diversity R code

Bernhard Klingenberg

1. Load Artist Datafile

```
path <- 'https://raw.githubusercontent.com/artofstat/ArtistDiversity/master/artistdata.csv'
artists <- read.csv(path)
```

2. Overall Statistics

```
library(dplyr)
```

```
##
## 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
```

```
# Overall number of artists across all museums:
artists %>%  summarize(size=n())
```

```
##    size
## 1 10108
```

```
# Number of artists in each museum:
artists %>%  group_by(museum) %>% summarize(size=n())
```

```
## # A tibble: 18 x 2
##    museum                                  size
##    <fct>                                  <int>
##  1 Art Institute of Chicago                 405
##  2 Dallas Museum of Art                     605
##  3 Denver Art Museum                        733
##  4 Detroit Institute of Arts                627
##  5 High Museum of Art                       402
##  6 Los Angeles County Museum of Art         635
##  7 Metropolitan Museum of Art, New York, NY 669
##  8 Museum of Contemporary Art               419
##  9 Museum of Fine Art Boston                611
## 10 Museum of Fine Arts Houston              696
## 11 Museum of Modern Art                     376
## 12 National Gallery of Art                  374
## 13 Nelson-Atkins Museum of Art              570
## 14 Philadelphia Museum of Art               654
## 15 Rhode Island School of Design Museum     620
## 16 San Francisco Museum of Modern Art       531
## 17 Whitney Museum of American Art           513
## 18 Yale University Art Gallery              668
```

```r
#Overall unique number of artists, after removing duplicates:
artists.unique <- artists %>% distinct(artist, .keep_all = TRUE)
artists.unique %>%  summarize(size=n())
```

```
##   size
## 1 9188
```

```r
### Overall statistics
# Gender Distribution:
table(artists.unique$gender, useNA="always")
```

```
##
##   man woman  <NA>
##  7086  1025  1077
```

```r
round(prop.table(table(artists.unique$gender)),3)
```

```
##
##   man woman
## 0.874 0.126
```

```r
#overall gender score confidence interval:
prop.test(1025, 1025+7086, correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  1025 out of 1025 + 7086, null probability 0.5
## X-squared = 4529.1, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.1193170 0.1337799
## sample estimates:
##         p
## 0.1263716
```

```r
# Ethnicity Distribution:
table(artists.unique$ethnicity, useNA="always")
```

```
##
##    asian    black hispanic    other    white     <NA>
##      668       91      210      109     6315     1795
```

```r
round(prop.table(table(artists.unique$ethnicity)),3)
```

```
##
##    asian    black hispanic    other    white
##    0.090    0.012    0.028    0.015    0.854
```

```r
# Simultaneous Score Confidence Intervals:
nums <- unlist(table(artists.unique$ethnicity))
sapply(nums, function(x) prop.test(x, sum(nums), correct=FALSE, conf.level = 1-0.05/5)$conf.int)
```

```
##           asian       black   hispanic      other     white
## [1,] 0.08213049 0.009415734 0.02383561 0.01154363 0.8432962
## [2,] 0.09931561 0.016076726 0.03382059 0.01881394 0.8644414
```

```r
# Gender & Ethnicity Distribution:
table(artists.unique$gender, artists.unique$ethnicity, useNA="always")
```

```
##
##         asian black hispanic other white <NA>
##   man     510    62      173    61  5121 1159
##   woman    35    26       20    29   732  183
##   <NA>    123     3       17    19   462  453
```

```
round(addmargins(prop.table(table(artists.unique$gender, artists.unique$ethnicity))),3)
```

```
##
##         asian black hispanic other white   Sum
##   man   0.075 0.009    0.026 0.009 0.757 0.876
##   woman 0.005 0.004    0.003 0.004 0.108 0.124
##   Sum   0.081 0.013    0.029 0.013 0.865 1.000
```

```
# Geographical Region:
table(artists.unique$GEO3major, useNA="always")
```

```
##
##                          Africa            Asia and the Pacific
##                              29                             661
##                          Europe Latin America and the Caribbean
##                            3329                             162
##                   North America                      West Asia
##                            3376                               7
##                            <NA>
##                            1624
```

```
round(prop.table(table(artists.unique$GEO3major)),3)
```

```
##
##                          Africa            Asia and the Pacific
##                           0.004                           0.087
##                          Europe Latin America and the Caribbean
##                           0.440                           0.021
##                   North America                      West Asia
##                           0.446                           0.001
```

```
# Birth Decade
mean(artists.unique$year, na.rm=TRUE)
```

```
## [1] 1863.428
```

3. Museum Specific Analysis

3.1 Gender

```
genderdf <- artists %>% select(museum, gender) %>% group_by(museum) %>%
summarize(men=sum(gender=="man", na.rm=TRUE),
          women=sum(gender=="woman", na.rm=TRUE),
          total=men+women,
          prop.women=women/total,
          LB=prop.test(women,total, correct=FALSE, conf.level = 1-0.05/18)$conf.int[1],
          UB=prop.test(women,total, correct=FALSE, conf.level = 1-0.05/18)$conf.int[2]
          )
genderdf$padj <- NA
for (i in 1:18) {
  genderdf$padj[i] <- prop.test(c(genderdf$women[i], sum(genderdf$women[-i])), c(genderdf$total[i], sum
```

```
}
genderdf$sig <- genderdf$padj<0.05/18
genderdf
```

```
## # A tibble: 18 x 9
##    museum           men women total prop.women     LB    UB     padj sig
##    <fct>          <int> <int> <int>      <dbl>  <dbl> <dbl>    <dbl> <lgl>
##  1 Art Institute~   314    45   359     0.125  0.0820 0.187 8.93e- 1 FALSE
##  2 Dallas Museum~   468    83   551     0.151  0.111  0.202 9.54e- 2 FALSE
##  3 Denver Art Mu~   585    90   675     0.133  0.0990 0.177 6.46e- 1 FALSE
##  4 Detroit Insti~   535    43   578     0.0744 0.0478 0.114 7.28e- 5 TRUE
##  5 High Museum o~   341    41   382     0.107  0.0686 0.164 2.24e- 1 FALSE
##  6 Los Angeles C~   490    58   548     0.106  0.0727 0.152 1.14e- 1 FALSE
##  7 Metropolitan ~   546    43   589     0.0730 0.0469 0.112 3.93e- 5 TRUE
##  8 Museum of Con~   292    97   389     0.249  0.190  0.320 1.94e-13 TRUE
##  9 Museum of Fin~   462    41   503     0.0815 0.0519 0.126 1.41e- 3 TRUE
## 10 Museum of Fin~   516    99   615     0.161  0.122  0.210 1.03e- 2 FALSE
## 11 Museum of Mod~   300    37   337     0.110  0.0686 0.171 3.16e- 1 FALSE
## 12 National Gall~   301    35   336     0.104  0.0642 0.165 1.88e- 1 FALSE
## 13 Nelson-Atkins~   450    59   509     0.116  0.0800 0.165 4.14e- 1 FALSE
## 14 Philadelphia ~   511    49   560     0.0875 0.0580 0.130 3.27e- 3 FALSE
## 15 Rhode Island ~   465    70   535     0.131  0.0932 0.181 8.20e- 1 FALSE
## 16 San Francisco~   404    89   493     0.181  0.135  0.238 2.97e- 4 TRUE
## 17 Whitney Museu~   367   104   471     0.221  0.169  0.283 4.90e-10 TRUE
## 18 Yale Universi~   518    68   586     0.116  0.0821 0.162 3.83e- 1 FALSE
```

3.2 Ethnicity

```
ethndf <- artists %>% select(museum, ethnicity) %>% group_by(museum) %>%
  summarize(asian=sum(ethnicity=="asian", na.rm=TRUE),
            black=sum(ethnicity=="black", na.rm=TRUE),
            hispanic=sum(ethnicity=="hispanic", na.rm=TRUE),
            other=sum(ethnicity=="other", na.rm=TRUE),
            white=sum(ethnicity=="white", na.rm=TRUE),
            total=asian+black+hispanic+other+white
            )
ethndf
```

```
## # A tibble: 18 x 7
##    museum                           asian black hispanic other white total
##    <fct>                            <int> <int>    <int> <int> <int> <int>
##  1 Art Institute of Chicago            24     1        7     1   309   342
##  2 Dallas Museum of Art                21     4       14    17   439   495
##  3 Denver Art Museum                   58     9       33    23   487   610
##  4 Detroit Institute of Arts           14     8        2     3   482   509
##  5 High Museum of Art                   3    37        5     3   300   348
##  6 Los Angeles County Museum of Art    91     0       15     6   401   513
##  7 Metropolitan Museum of Art, New~    43     1        8     7   472   531
##  8 Museum of Contemporary Art          26    10       24     5   312   377
##  9 Museum of Fine Art Boston           76     5       10     4   377   472
## 10 Museum of Fine Arts Houston         24     6       27     7   496   560
## 11 Museum of Modern Art                30     6       11     4   249   300
## 12 National Gallery of Art              4     0        2     2   301   309
## 13 Nelson-Atkins Museum of Art         45     2        6    11   408   472
```

```
## 14 Philadelphia Museum of Art          44     6       13     2   468   533
## 15 Rhode Island School of Design M~    72     5       15    12   374   478
## 16 San Francisco Museum of Modern ~    32     9       15     5   389   450
## 17 Whitney Museum of American Art      12    10       10     4   398   434
## 18 Yale University Art Gallery         80     4       13     6   460   563
```

3.2.1 Asian

```
ethndf.asian <- ethndf %>% select(museum, asian, total) %>% group_by(museum) %>%
  mutate(prop=asian/total,
         LB=prop.test(asian, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[1],
         UB=prop.test(asian, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[2]
         )
ethndf.asian$padj <- NA
for (i in 1:18) {
  ethndf.asian$padj[i] <- prop.test(c(ethndf.asian$asian[i], sum(ethndf.asian$asian[-i])), c(ethndf.asia
}
ethndf.asian$sig <- ethndf.asian$padj<0.05/(18*5)
ethndf.asian
```

```
## # A tibble: 18 x 8
## # Groups:   museum [18]
##    museum                asian total    prop      LB     UB     padj sig
##    <fct>                 <int> <int>   <dbl>   <dbl>  <dbl>    <dbl> <lgl>
##  1 Art Institute of Chi~    24   342 0.0702  0.0356  0.134  3.38e- 1 FALSE
##  2 Dallas Museum of Art     21   495 0.0424  0.0205  0.0859 5.50e- 4 TRUE
##  3 Denver Art Museum        58   610 0.0951  0.0615  0.144  3.17e- 1 FALSE
##  4 Detroit Institute of~    14   509 0.0275  0.0113  0.0653 1.96e- 6 TRUE
##  5 High Museum of Art        3   348 0.00862 0.00149 0.0483 2.11e- 7 TRUE
##  6 Los Angeles County M~    91   513 0.177   0.127   0.243  4.50e-15 TRUE
##  7 Metropolitan Museum ~    43   531 0.0810  0.0487  0.132  7.79e- 1 FALSE
##  8 Museum of Contempora~    26   377 0.0690  0.0359  0.128  2.74e- 1 FALSE
##  9 Museum of Fine Art B~    76   472 0.161   0.111   0.228  6.33e-10 TRUE
## 10 Museum of Fine Arts ~    24   560 0.0429  0.0216  0.0831 2.60e- 4 TRUE
## 11 Museum of Modern Art     30   300 0.1     0.0547  0.176  3.17e- 1 FALSE
## 12 National Gallery of ~     4   309 0.0129  0.00272 0.0594 4.24e- 6 TRUE
## 13 Nelson-Atkins Museum~    45   472 0.0953  0.0581  0.152  3.72e- 1 FALSE
## 14 Philadelphia Museum ~    44   533 0.0826  0.0500  0.133  8.83e- 1 FALSE
## 15 Rhode Island School ~    72   478 0.151   0.103   0.216  7.40e- 8 TRUE
## 16 San Francisco Museum~    32   450 0.0711  0.0394  0.125  3.02e- 1 FALSE
## 17 Whitney Museum of Am~    12   434 0.0276  0.0106  0.0699 1.29e- 5 TRUE
## 18 Yale University Art ~    80   563 0.142   0.0987  0.200  3.10e- 7 TRUE
```

3.2.2 Black

```
ethndf.black <- ethndf %>% select(museum, black, total) %>% group_by(museum) %>%
  mutate(prop=black/total,
         LB=prop.test(black, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[1],
         UB=prop.test(black, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[2]
         )
ethndf.black$padj <- NA
for (i in 1:18) {
  ethndf.black$padj[i] <- prop.test(c(ethndf.black$black[i], sum(ethndf.black$black[-i])), c(ethndf.bla
}
```

```
## Warning in prop.test(c(ethndf.black$black[i], sum(ethndf.black$black[-
## i])), : Chi-squared approximation may be incorrect

## Warning in prop.test(c(ethndf.black$black[i], sum(ethndf.black$black[-
## i])), : Chi-squared approximation may be incorrect
```

```
ethndf.black$sig <- ethndf.black$padj<0.05/(18*5)
ethndf.black
```

```
## # A tibble: 18 x 8
## # Groups:   museum [18]
##    museum                black total    prop      LB     UB     padj sig
##    <fct>                 <int> <int>   <dbl>   <dbl>  <dbl>    <dbl> <lgl>
##  1 Art Institute of Chi~     1   342 0.00292 2.11e-4 0.0391 6.29e- 2 FALSE
##  2 Dallas Museum of Art      4   495 0.00808 1.70e-3 0.0376 2.00e- 1 FALSE
##  3 Denver Art Museum         9   610 0.0148  4.95e-3 0.0432 9.88e- 1 FALSE
##  4 Detroit Institute of~     8   509 0.0157  4.96e-3 0.0486 8.64e- 1 FALSE
##  5 High Museum of Art       37   348 0.106   6.18e-2 0.177  3.43e-47 TRUE
##  6 Los Angeles County M~     0   513 0       0.      0.0227 4.12e- 3 FALSE
##  7 Metropolitan Museum ~     1   531 0.00188 1.36e-4 0.0255 1.07e- 2 FALSE
##  8 Museum of Contempora~    10   377 0.0265  9.38e-3 0.0727 5.44e- 2 FALSE
##  9 Museum of Fine Art B~     5   472 0.0106  2.56e-3 0.0427 4.33e- 1 FALSE
## 10 Museum of Fine Arts ~     6   560 0.0107  2.89e-3 0.0389 4.04e- 1 FALSE
## 11 Museum of Modern Art      6   300 0.02    5.40e-3 0.0713 4.50e- 1 FALSE
## 12 National Gallery of ~     0   309 0       0.      0.0371 2.80e- 2 FALSE
## 13 Nelson-Atkins Museum~     2   472 0.00424 5.41e-4 0.0324 5.00e- 2 FALSE
## 14 Philadelphia Museum ~     6   533 0.0113  3.03e-3 0.0409 4.81e- 1 FALSE
## 15 Rhode Island School ~     5   478 0.0105  2.53e-3 0.0422 4.16e- 1 FALSE
## 16 San Francisco Museum~     9   450 0.02    6.71e-3 0.0581 3.50e- 1 FALSE
## 17 Whitney Museum of Am~    10   434 0.0230  8.15e-3 0.0634 1.46e- 1 FALSE
## 18 Yale University Art ~     4   563 0.00710 1.49e-3 0.0332 1.16e- 1 FALSE
```

3.2.3 Hispanic

```
ethndf.hispanic <- ethndf %>% select(museum, hispanic, total) %>% group_by(museum) %>%
  mutate(prop=hispanic/total,
         LB=prop.test(hispanic, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[1],
         UB=prop.test(hispanic, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[2]
         )
ethndf.hispanic$padj <- NA
for (i in 1:18) {
  ethndf.hispanic$padj[i] <- prop.test(c(ethndf.hispanic$hispanic[i], sum(ethndf.hispanic$hispanic[-i]))
}
ethndf.hispanic$sig <- ethndf.hispanic$padj<0.05/(18*5)
ethndf.hispanic
```

```
## # A tibble: 18 x 8
## # Groups:   museum [18]
##    museum              hispanic total    prop       LB     UB    padj sig
##    <fct>                  <int> <int>   <dbl>    <dbl>  <dbl>   <dbl> <lgl>
##  1 Art Institute of ~         7   342 0.0205  0.00602  0.0672 4.04e-1 FALSE
##  2 Dallas Museum of ~        14   495 0.0283  0.0116   0.0671 9.38e-1 FALSE
##  3 Denver Art Museum         33   610 0.0541  0.0302   0.0951 3.76e-5 TRUE
##  4 Detroit Institute~         2   509 0.00393 0.000502 0.0301 7.38e-4 FALSE
##  5 High Museum of Art         5   348 0.0144  0.00348  0.0574 1.21e-1 FALSE
```

```
##  6 Los Angeles Count~       15   513 0.0292  0.0124   0.0675 8.29e-1 FALSE
##  7 Metropolitan Muse~        8   531 0.0151  0.00476  0.0467 6.63e-2 FALSE
##  8 Museum of Contemp~       24   377 0.0637  0.0322   0.122  1.36e-5 TRUE
##  9 Museum of Fine Ar~       10   472 0.0212  0.00749  0.0585 3.73e-1 FALSE
## 10 Museum of Fine Ar~       27   560 0.0482  0.0253   0.0900 2.23e-3 FALSE
## 11 Museum of Modern ~       11   300 0.0367  0.0136   0.0952 3.37e-1 FALSE
## 12 National Gallery ~        2   309 0.00647 0.000827 0.0488 2.04e-2 FALSE
## 13 Nelson-Atkins Mus~        6   472 0.0127  0.00343  0.0460 4.08e-2 FALSE
## 14 Philadelphia Muse~       13   533 0.0244  0.00972  0.0599 6.28e-1 FALSE
## 15 Rhode Island Scho~       15   478 0.0314  0.0133   0.0723 6.16e-1 FALSE
## 16 San Francisco Mus~       15   450 0.0333  0.0141   0.0766 4.56e-1 FALSE
## 17 Whitney Museum of~       10   434 0.0230  0.00815  0.0634 5.42e-1 FALSE
## 18 Yale University A~       13   563 0.0231  0.00920  0.0568 4.88e-1 FALSE
```

3.2.4 White

```
ethndf.white <- ethndf %>% select(museum, white, total) %>% group_by(museum) %>%
  mutate(prop=white/total,
         LB=prop.test(white, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[1],
         UB=prop.test(white, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[2]
         )
ethndf.white$padj <- NA
for (i in 1:18) {
  ethndf.white$padj[i] <- prop.test(c(ethndf.white$white[i], sum(ethndf.white$white[-i])), c(ethndf.whi
}
ethndf.white$sig <- ethndf.white$padj<0.05/(18*5)
ethndf.white
```

```
## # A tibble: 18 x 8
## # Groups:   museum [18]
##    museum                   white total  prop    LB    UB       padj sig
##    <fct>                    <int> <int> <dbl> <dbl> <dbl>      <dbl> <lgl>
##  1 Art Institute of Chicago   309   342 0.904 0.834 0.946    1.47e-2 FALSE
##  2 Dallas Museum of Art       439   495 0.887 0.828 0.927    6.17e-2 FALSE
##  3 Denver Art Museum          487   610 0.798 0.737 0.848    9.59e-6 TRUE
##  4 Detroit Institute of Ar~   482   509 0.947 0.901 0.972    3.41e-9 TRUE
##  5 High Museum of Art         300   348 0.862 0.786 0.914    8.45e-1 FALSE
##  6 Los Angeles County Muse~   401   513 0.782 0.713 0.838    2.56e-7 TRUE
##  7 Metropolitan Museum of ~   472   531 0.889 0.833 0.928    3.77e-2 FALSE
##  8 Museum of Contemporary ~   312   377 0.828 0.751 0.884    7.81e-2 FALSE
##  9 Museum of Fine Art Bost~   377   472 0.799 0.728 0.855    1.25e-4 TRUE
## 10 Museum of Fine Arts Hou~   496   560 0.886 0.831 0.924    5.56e-2 FALSE
## 11 Museum of Modern Art       249   300 0.83  0.743 0.892    1.49e-1 FALSE
## 12 National Gallery of Art    301   309 0.974 0.921 0.992    2.80e-9 TRUE
## 13 Nelson-Atkins Museum of~   408   472 0.864 0.801 0.910    7.04e-1 FALSE
## 14 Philadelphia Museum of ~   468   533 0.878 0.821 0.919    1.80e-1 FALSE
## 15 Rhode Island School of ~   374   478 0.782 0.711 0.840    8.90e-7 TRUE
## 16 San Francisco Museum of~   389   450 0.864 0.799 0.911    7.09e-1 FALSE
## 17 Whitney Museum of Ameri~   398   434 0.917 0.859 0.952    3.24e-4 TRUE
## 18 Yale University Art Gal~   460   563 0.817 0.754 0.867    3.48e-3 FALSE
```

3.2.5 Other

```
ethndf.other <- ethndf %>% select(museum, other, total) %>% group_by(museum) %>%
  mutate(prop=other/total,
         LB=prop.test(other, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[1],
         UB=prop.test(other, total, correct=FALSE, conf.level = 1-0.05/(18*5))$conf.int[2]
         )
ethndf.other$padj <- NA
for (i in 1:18) {
  ethndf.other$padj[i] <- prop.test(c(ethndf.other$other[i], sum(ethndf.other$other[-i])), c(ethndf.othe
}
```

```
## Warning in prop.test(c(ethndf.other$other[i], sum(ethndf.other$other[-
## i])), : Chi-squared approximation may be incorrect
```

```
## Warning in prop.test(c(ethndf.other$other[i], sum(ethndf.other$other[-
## i])), : Chi-squared approximation may be incorrect
```

```
ethndf.other$sig <- ethndf.other$padj<0.05/(18*5)
ethndf.other
```

```
## # A tibble: 18 x 8
## # Groups:   museum [18]
##    museum                other total    prop       LB     UB     padj sig
##    <fct>                 <int> <int>   <dbl>    <dbl>  <dbl>    <dbl> <lgl>
##  1 Art Institute of Ch~      1   342 0.00292 0.000211 0.0391  6.45e-2 FALSE
##  2 Dallas Museum of Art     17   495 0.0343  0.0153   0.0753  1.82e-4 TRUE
##  3 Denver Art Museum        23   610 0.0377  0.0187   0.0744  9.46e-7 TRUE
##  4 Detroit Institute o~      3   509 0.00589 0.00102  0.0334  8.82e-2 FALSE
##  5 High Museum of Art        3   348 0.00862 0.00149  0.0483  3.35e-1 FALSE
##  6 Los Angeles County ~      6   513 0.0117  0.00315  0.0424  5.59e-1 FALSE
##  7 Metropolitan Museum~      7   531 0.0132  0.00387  0.0439  7.63e-1 FALSE
##  8 Museum of Contempor~      5   377 0.0133  0.00321  0.0532  8.12e-1 FALSE
##  9 Museum of Fine Art ~      4   472 0.00847 0.00178  0.0394  2.47e-1 FALSE
## 10 Museum of Fine Arts~      7   560 0.0125  0.00367  0.0416  6.53e-1 FALSE
## 11 Museum of Modern Art      4   300 0.0133  0.00280  0.0611  8.41e-1 FALSE
## 12 National Gallery of~      2   309 0.00647 0.000827 0.0488  2.20e-1 FALSE
## 13 Nelson-Atkins Museu~     11   472 0.0233  0.00862  0.0615  1.10e-1 FALSE
## 14 Philadelphia Museum~      2   533 0.00375 0.000479 0.0287  2.99e-2 FALSE
## 15 Rhode Island School~     12   478 0.0251  0.00966  0.0637  5.17e-2 FALSE
## 16 San Francisco Museu~      5   450 0.0111  0.00269  0.0448  5.15e-1 FALSE
## 17 Whitney Museum of A~      4   434 0.00922 0.00193  0.0427  3.29e-1 FALSE
## 18 Yale University Art~      6   563 0.0107  0.00287  0.0387  4.08e-1 FALSE
```

3.3 Geographic Origin

```
geodf <- artists %>% select(museum, GEO3major) %>% group_by(museum) %>%
  summarize(Africa=round(100*prop.table(table(GEO3major))[1],1),
            Asia=round(100*prop.table(table(GEO3major))[2],1),
            Europe=round(100*prop.table(table(GEO3major))[3],1),
            LatAm=round(100*prop.table(table(GEO3major))[4],1),
            NorthAm=round(100*prop.table(table(GEO3major))[5],1),
            WAsia=round(100*prop.table(table(GEO3major))[6],1)
  )
geodf
```

```
## # A tibble: 18 x 7
##    museum                          Africa  Asia Europe LatAm NorthAm WAsia
##    <fct>                            <dbl> <dbl>  <dbl> <dbl>   <dbl> <dbl>
##  1 Art Institute of Chicago           0     6.5   56.5   1.4    35.5   0
##  2 Dallas Museum of Art               0.2   4.4   45.1   1.4    48.9   0
##  3 Denver Art Museum                  0.5   8.4   29.7   3.1    58.1   0.3
##  4 Detroit Institute of Arts          0.2   2.9   59.5   0.6    36.9   0
##  5 High Museum of Art                 2.5   0.3   37.8   0.8    58.6   0
##  6 Los Angeles County Museum of A~    0.4  17.4   44.4   2.4    35.5   0
##  7 Metropolitan Museum of Art, Ne~    0.2   9.5   63.6   0.8    25.7   0.2
##  8 Museum of Contemporary Art         0.5   5.9   22.3   4      67.3   0
##  9 Museum of Fine Art Boston          0    16.3   51.2   1.9    30.6   0
## 10 Museum of Fine Arts Houston        0.4   4.4   38.6   4      52.5   0.2
## 11 Museum of Modern Art               1    10.5   47.6   3.1    37.8   0
## 12 National Gallery of Art            0     0.9   56.9   0      42.2   0
## 13 Nelson-Atkins Museum of Art        0     9.7   37.4   0.9    51.8   0.2
## 14 Philadelphia Museum of Art         0.4   7.5   61.9   1.9    28.3   0
## 15 Rhode Island School of Design ~    0    13.5   44.2   3.6    38.6   0.2
## 16 San Francisco Museum of Modern~    1.3   7.2   32.8   3.8    55     0
## 17 Whitney Museum of American Art     0     2.1   11.1   1.9    84.7   0.2
## 18 Yale University Art Gallery        0    14.1   39.7   1.9    44.1   0.2
```

3.4 Birth Year

```
yeardf <- artists %>% select(museum, year) %>% group_by(museum) %>%
  summarize(Avg.Year=round(mean(year, na.rm=TRUE)))
yeardf
```

```
## # A tibble: 18 x 2
##    museum                                  Avg.Year
##    <fct>                                      <dbl>
##  1 Art Institute of Chicago                    1836
##  2 Dallas Museum of Art                        1886
##  3 Denver Art Museum                           1886
##  4 Detroit Institute of Arts                   1802
##  5 High Museum of Art                          1866
##  6 Los Angeles County Museum of Art            1885
##  7 Metropolitan Museum of Art, New York, NY    1804
##  8 Museum of Contemporary Art                  1949
##  9 Museum of Fine Art Boston                   1803
## 10 Museum of Fine Arts Houston                 1891
## 11 Museum of Modern Art                        1921
## 12 National Gallery of Art                     1813
## 13 Nelson-Atkins Museum of Art                 1850
## 14 Philadelphia Museum of Art                  1806
## 15 Rhode Island School of Design Museum        1849
## 16 San Francisco Museum of Modern Art          1929
## 17 Whitney Museum of American Art              1932
## 18 Yale University Art Gallery                 1851
```