

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
```

```

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

##      size
## 1 9188

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

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

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

##
##   man woman
## 0.874 0.126

#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

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

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

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

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

# 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

# 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
```

```
## # A tibble: 18 x 8
##   museum      men women total prop.women    LB    UB    padj
##   <fct>      <int> <int> <int>      <dbl> <dbl> <dbl> <dbl>
## 1 Art Institute of Ch~ 314   45  359    0.125 0.0820 0.187 1.61e+ 1
## 2 Dallas Museum of Art 468   83  551    0.151 0.111 0.202 1.72e+ 0
## 3 Denver Art Museum   585   90  675    0.133 0.0990 0.177 1.16e+ 1
## 4 Detroit Institute o~ 535   43  578    0.0744 0.0478 0.114 1.31e- 3
## 5 High Museum of Art   341   41  382    0.107 0.0686 0.164 4.03e+ 0
## 6 Los Angeles County ~ 490   58  548    0.106 0.0727 0.152 2.06e+ 0
## 7 Metropolitan Museum~ 546   43  589    0.0730 0.0469 0.112 7.08e- 4
## 8 Museum of Contempor~ 292   97  389    0.249 0.190 0.320 3.49e-12
## 9 Museum of Fine Art ~ 462   41  503    0.0815 0.0519 0.126 2.54e- 2
## 10 Museum of Fine Arts~ 516   99  615    0.161 0.122 0.210 1.86e- 1
## 11 Museum of Modern Art 300   37  337    0.110 0.0686 0.171 5.69e+ 0
## 12 National Gallery of~ 301   35  336    0.104 0.0642 0.165 3.39e+ 0
## 13 Nelson-Atkins Museu~ 450   59  509    0.116 0.0800 0.165 7.44e+ 0
## 14 Philadelphia Museum~ 511   49  560    0.0875 0.0580 0.130 5.89e- 2
## 15 Rhode Island School~ 465   70  535    0.131 0.0932 0.181 1.48e+ 1
## 16 San Francisco Museu~ 404   89  493    0.181 0.135 0.238 5.35e- 3
## 17 Whitney Museum of A~ 367  104  471    0.221 0.169 0.283 8.82e- 9
## 18 Yale University Art~ 518   68  586    0.116 0.0821 0.162 6.90e+ 0
```

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.asian$total[i], sum(ethndf.asian$total[-i])))$p.value
}
ethndf.asian
```

```
## # A tibble: 18 x 7
## # Groups:   museum [18]
##   museum      asian total      prop      LB      UB      padj
##   <fct>      <int> <int>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Art Institute of Chicago      24   342 0.0702  0.0356  0.134  3.04e+ 1
## 2 Dallas Museum of Art         21   495 0.0424  0.0205  0.0859  4.95e- 2
## 3 Denver Art Museum           58   610 0.0951  0.0615  0.144  2.86e+ 1
## 4 Detroit Institute of Arts     14   509 0.0275  0.0113  0.0653  1.76e- 4
## 5 High Museum of Art            3   348 0.00862 0.00149 0.0483  1.90e- 5
## 6 Los Angeles County Museum ~   91   513 0.177   0.127   0.243  4.05e-13
## 7 Metropolitan Museum of Art~   43   531 0.0810  0.0487  0.132  7.01e+ 1
## 8 Museum of Contemporary Art    26   377 0.0690  0.0359  0.128  2.47e+ 1
## 9 Museum of Fine Art Boston     76   472 0.161   0.111   0.228  5.69e- 8
## 10 Museum of Fine Arts Houston  24   560 0.0429  0.0216  0.0831  2.34e- 2
## 11 Museum of Modern Art        30   300 0.1     0.0547  0.176  2.86e+ 1
## 12 National Gallery of Art       4   309 0.0129  0.00272 0.0594  3.81e- 4
## 13 Nelson-Atkins Museum of Art   45   472 0.0953  0.0581  0.152  3.35e+ 1
## 14 Philadelphia Museum of Art   44   533 0.0826  0.0500  0.133  7.95e+ 1
## 15 Rhode Island School of Des~  72   478 0.151   0.103   0.216  6.66e- 6
## 16 San Francisco Museum of Mo~  32   450 0.0711  0.0394  0.125  2.72e+ 1
## 17 Whitney Museum of American~  12   434 0.0276  0.0106  0.0699  1.16e- 3
## 18 Yale University Art Gallery  80   563 0.142   0.0987  0.200  2.79e- 5
```

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.black$total[i], sum(ethndf.black$total[-i])))$p.value
}
```

```
## 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
```

```
## # A tibble: 18 x 7
## # Groups:   museum [18]
##   museum          black total    prop    LB    UB    padj
##   <fct>          <int> <int>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Art Institute of Chicago      1   342 0.00292 2.11e-4 0.0391 5.66e+ 0
## 2 Dallas Museum of Art         4   495 0.00808 1.70e-3 0.0376 1.80e+ 1
## 3 Denver Art Museum           9   610 0.0148 4.95e-3 0.0432 8.89e+ 1
## 4 Detroit Institute of Arts     8   509 0.0157 4.96e-3 0.0486 7.77e+ 1
## 5 High Museum of Art          37   348 0.106 6.18e-2 0.177 3.08e-45
## 6 Los Angeles County Museum ~  0   513 0      0.      0.0227 3.71e- 1
## 7 Metropolitan Museum of Art~  1   531 0.00188 1.36e-4 0.0255 9.67e- 1
## 8 Museum of Contemporary Art    10   377 0.0265 9.38e-3 0.0727 4.90e+ 0
## 9 Museum of Fine Art Boston     5   472 0.0106 2.56e-3 0.0427 3.90e+ 1
## 10 Museum of Fine Arts Houston  6   560 0.0107 2.89e-3 0.0389 3.64e+ 1
## 11 Museum of Modern Art         6   300 0.02    5.40e-3 0.0713 4.05e+ 1
## 12 National Gallery of Art      0   309 0      0.      0.0371 2.52e+ 0
## 13 Nelson-Atkins Museum of Art  2   472 0.00424 5.41e-4 0.0324 4.50e+ 0
## 14 Philadelphia Museum of Art   6   533 0.0113 3.03e-3 0.0409 4.33e+ 1
## 15 Rhode Island School of Des~  5   478 0.0105 2.53e-3 0.0422 3.74e+ 1
## 16 San Francisco Museum of Mo~  9   450 0.02    6.71e-3 0.0581 3.15e+ 1
## 17 Whitney Museum of American~ 10   434 0.0230 8.15e-3 0.0634 1.31e+ 1
## 18 Yale University Art Gallery  4   563 0.00710 1.49e-3 0.0332 1.05e+ 1
```

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
```

```
## # A tibble: 18 x 7
## # Groups:   museum [18]
##   museum          hispanic total    prop    LB    UB    padj
##   <fct>          <int> <int>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Art Institute of Chicago      7   342 0.0205 0.00602 0.0672 3.63e+1
## 2 Dallas Museum of Art        14   495 0.0283 0.0116 0.0671 8.44e+1
## 3 Denver Art Museum          33   610 0.0541 0.0302 0.0951 3.38e-3
## 4 Detroit Institute of Ar~     2   509 0.00393 0.000502 0.0301 6.64e-2
## 5 High Museum of Art          5   348 0.0144 0.00348 0.0574 1.09e+1
## 6 Los Angeles County Muse~    15   513 0.0292 0.0124 0.0675 7.46e+1
## 7 Metropolitan Museum of ~     8   531 0.0151 0.00476 0.0467 5.97e+0
## 8 Museum of Contemporary ~    24   377 0.0637 0.0322 0.122 1.23e-3
```

```
## 9 Museum of Fine Art Bost~ 10 472 0.0212 0.00749 0.0585 3.36e+1
## 10 Museum of Fine Arts Hou~ 27 560 0.0482 0.0253 0.0900 2.00e-1
## 11 Museum of Modern Art 11 300 0.0367 0.0136 0.0952 3.03e+1
## 12 National Gallery of Art 2 309 0.00647 0.000827 0.0488 1.84e+0
## 13 Nelson-Atkins Museum of~ 6 472 0.0127 0.00343 0.0460 3.67e+0
## 14 Philadelphia Museum of ~ 13 533 0.0244 0.00972 0.0599 5.65e+1
## 15 Rhode Island School of ~ 15 478 0.0314 0.0133 0.0723 5.54e+1
## 16 San Francisco Museum of~ 15 450 0.0333 0.0141 0.0766 4.11e+1
## 17 Whitney Museum of Ameri~ 10 434 0.0230 0.00815 0.0634 4.87e+1
## 18 Yale University Art Gal~ 13 563 0.0231 0.00920 0.0568 4.39e+1
```

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
```

```
## # A tibble: 18 x 7
## # Groups:   museum [18]
##   museum      white total  prop    LB    UB      padj
##   <fct>      <int> <int> <dbl> <dbl> <dbl>    <dbl>
## 1 Art Institute of Chicago      309    342 0.904 0.834 0.946    1.32e+0
## 2 Dallas Museum of Art          439    495 0.887 0.828 0.927    5.55e+0
## 3 Denver Art Museum            487    610 0.798 0.737 0.848    8.63e-4
## 4 Detroit Institute of Arts     482    509 0.947 0.901 0.972    3.07e-7
## 5 High Museum of Art           300    348 0.862 0.786 0.914    7.60e+1
## 6 Los Angeles County Museum of ~ 401    513 0.782 0.713 0.838    2.31e-5
## 7 Metropolitan Museum of Art, N~ 472    531 0.889 0.833 0.928    3.40e+0
## 8 Museum of Contemporary Art     312    377 0.828 0.751 0.884    7.03e+0
## 9 Museum of Fine Art Boston      377    472 0.799 0.728 0.855    1.13e-2
## 10 Museum of Fine Arts Houston    496    560 0.886 0.831 0.924    5.00e+0
## 11 Museum of Modern Art          249    300 0.83  0.743 0.892    1.34e+1
## 12 National Gallery of Art       301    309 0.974 0.921 0.992    2.52e-7
## 13 Nelson-Atkins Museum of Art    408    472 0.864 0.801 0.910    6.34e+1
## 14 Philadelphia Museum of Art     468    533 0.878 0.821 0.919    1.62e+1
## 15 Rhode Island School of Design~ 374    478 0.782 0.711 0.840    8.01e-5
## 16 San Francisco Museum of Moder~ 389    450 0.864 0.799 0.911    6.38e+1
## 17 Whitney Museum of American Art 398    434 0.917 0.859 0.952    2.91e-2
## 18 Yale University Art Gallery    460    563 0.817 0.754 0.867    3.13e-1
```

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]
  )
```


## 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
```