Rowan Gray

Prof. Cannon

CS36

11 December 2024

Final Project: Analysis of the MET Collection

Introduction

This project will analyze data from the Metropolitan Museum of Art in New York that describes their collection. This dataset contains a variable called `Is Highlight`, which according to the MET's website, is a boolean variable set to true if an artwork is "a popular and important artwork in the collection". I'm really interested to see what variables predict an artwork's status as a popular and important work and seeing what trends I can find. Here's three specific questions on that theme.

Question 1: Which variables have the highest correlation with `isHighlight`? Question 2: If I'm artist and want the MET to think my art is "a popular and important artwork in the collection" what kind of art should I make? Question 3: What time periods and cultures are the best-represented among the MET's "popular and important artwork"?

I answer these questions in much more robust detail in the Conclusions section, but here's a topline summary of my results.

1. The variables with the largest and most robust effect on whether an artwork is a highlight have to do with the medium. Specifically, Stone Sculpture and Paintings are the most successful.
2. Culture seems to have no impact on the highlight worthiness of your art most of the time, but there are some edge cases.
3. What year you created your work has no impact on whether your art is a highlight, but what year the MET bought your work is extremely robustly associated with highlight status, but at a very low level.

The code for this project is opensourced on Github. You can find it here.

Data Source

The data for this project was sourced from The Metropolitan Museum of Art Open Access CSV. This is a dataset of roughly 470,000 artworks in the MET's collection, on display and in their warehouses. On the MET's repo for this data, they say that it is "generated from our internal database". The dataset was published under a Creative Commons Zero license and the MET all copyright and related rights to the dataset. The MET has an incentive to have an accurate database of all the art in their collection and to release it to the public, so the dataset is likely as high-quality as the MET's data scientists can make it. That said, I don't think the dataset is natively stored in a CSV. There were some issues parsing data, especially gender data, that seem related to an imperfect translation to CSV. Also most big datasets aren't usually stored in CSV, it's just a convenient format to store your data in if you want to opensource it.

Data Ethics

The project has several ethical concerns.

I analyzed the Culture variable in the dataset, but doing this required that I both tidy and flatten the data. As a result, nearly all sub-cultures were removed as independent classifications, and I only kept the top 20 or so largest cultures- all cultures with more than 1000 pieces in the museum. This should not be taken to be an accurate and full depiction of the cultural diversity of objects at the MET museum, since all small cultures and most subcultures were removed. For context, the dataset originally had a total of around 7300 cultures, and my analysis flattened it down to 23.

Additionally, the way I parsed the gender variable right now marks everyone who has a gender that isn't explicitly marked Female as Male. I looked through the data briefly and couldn't find any cases of this messing up but I'm sure it does sometimes. This also posits Male as implicitly the default gender which makes this method of data cleaning problematic, especially going forward, but it seems pretty accurate here because the MET's collection is so Male dominated. And of course there's lots of cases where the art was made by more than one person or by a company.

Finally, I'm concerned that this project could be taken as an attempt to figure out what kind of art is the 'most valuable'. This would both be an inaccurate reading of what I'm trying to do and of what I think about art in general. I'm trying to figure out what the MET thinks is highlight-worthy art and what, if any, variables in their publicly released dataset predict that. Art in itself does not have any value—it only exists relative to the people who made it and the people who engage with it.

Data Import, Cleaning and Tidying, and Exploration

Data Import

First things first, let's

Import our libraries-

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(splines)
library(modelr)
```

And load our data-

```
met_objects <- read_csv("openaccess/MetObjects.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 484956 Columns: 54
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (45): Object Number, Gallery Number, Department, Object Name, Title, Cul...
## dbl  (5): Object ID, AccessionYear, Constituent ID, Object Begin Date, Objec...
## lgl  (4): Is Highlight, Is Timeline Work, Is Public Domain, Metadata Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Cleaning and Tidying

Variables that seem like they could possibly be useful for predicting `Is Highlight`

1. Department
2. AccessionYear
3. Culture
4. Artist End Date
5. Artist Gender
6. Object Date
7. Classification

I suspect that some of these might be categorical variables with a lot of unique values that we should collapse. Let's test that theory by making some data grids

```
unique_department <- data_grid(met_objects, Department)
unique_culture <- data_grid(met_objects, Culture)
unique_classification <- data_grid(met_objects, Classification)
```

Looking the sizes of the data grids using our sidebar, we're going to need to figure out how to collapse unique values in everything except `Department`.

Let's start with culture!

Here's some code that collapses this a bunch- down to about 3,000 unique cultures from 7,300 or so originally.

```
unique_culture[1, "Culture"] <- "Chugach, Native American"
unique_culture <- unique_culture %>%
  mutate(Culture_Group = str_extract(Culture, "^[^[,;(]]+")) %>%
  mutate(Culture_Group = ifelse(
    str_detect(Culture_Group, "^(possibly|probably)"),
    str_extract(Culture_Group, "(?<=(possibly|probably)\\s)\\w+"),
    Culture_Group
  )) %>%
  mutate(Culture_Group = str_to_title(Culture_Group))
```

Now let's merge this with our original dataframe so that we can filter by top cultures.

```
met_objects <- met_objects %>%
  left_join(unique_culture, "Culture") %>%
  mutate(Culture_Group = as.factor(Culture_Group)) %>%
  mutate(Culture_Group = fct_lump_min(Culture_Group, min = 1000))

met_objects %>%
  select(Culture_Group) %>%
  group_by(Culture_Group) %>%
  count() %>%
  arrange(desc(n))
```

```
## # A tibble: 24 x 2
## # Groups:   Culture_Group [24]
##    Culture_Group      n
##    <fct>          <int>
##  1 <NA>          276768
##  2 Other          42313
##  3 American       29348
##  4 French         23836
##  5 Greek          20653
##  6 Japan          16941
##  7 China          13533
##  8 Italian        10679
##  9 British        10614
```

```
## 10 German          7793
## # i 14 more rows
```

Okay now let's do classification

It seems like we could filter a bunch by just taking the first word.

```
unique_classification <- unique_classification %>%
  mutate(Classification_Group = str_extract(Classification, "^[^[-,;(|]]+")) %>%
  mutate(Classification_Group = str_to_title(Classification_Group))
```

Making it a factor and taking only the top groups.

```
met_objects <- met_objects %>%
  left_join(unique_classification, "Classification") %>%
  mutate(Classification_Group = as.factor(Classification_Group)) %>%
  mutate(Classification_Group = fct_lump_min(Classification_Group, min = 1000))
```

Cleaning up object date

```
met_objects <- met_objects %>%
  mutate(
    parsed_year = `Object Date`,
    parsed_year = if_else(
      str_detect(parsed_year, "\\d{4}"),
      str_extract(parsed_year, "\\d{4}"),
      NA_character_
    ),
    parsed_year = if_else(
      is.na(parsed_year) & str_detect(`Object Date`, "\\d{3}"),
      str_extract(`Object Date`, "\\d{3}"),
      parsed_year
    ),
    parsed_year = if_else(
      is.na(parsed_year) & str_detect(`Object Date`, "\\d+\\s*(BCE|CE|BC)"),
      if_else(
        str_detect(`Object Date`, "BCE|BC"),
        as.character(-as.numeric(str_extract(`Object Date`, "\\d+"))),
        str_extract(`Object Date`, "\\d+")
      ),
      parsed_year
    ),
    parsed_year = if_else(
      is.na(parsed_year) & str_detect(`Object Date`, "\\d{1,2}(st|nd|rd|th)"),
      as.character(
        as.numeric(str_extract(`Object Date`, "\\d{1,2}")) * 100 - 50
      ),
      parsed_year
    ),
    parsed_year = as.numeric(parsed_year)
  )
```

Cleaning AccessionYear- much less work!

```
met_objects$AccessionYear <- as.double(substr(met_objects$AccessionYear, 1, 4))
```

Cleaning Gender

```
met_objects <- met_objects %>%
  mutate(parsed_gender = ifelse(is.na(`Artist Gender`), NA,
                                ifelse(str_detect(`Artist Gender`, "[Ff]emale"), "Female", "Male")))
```

Data Transformation/Visualization/Modeling

Messing with the data types to make glm happy and selecting just the variables we need

```
met_objects$Culture_Group <- as.character(met_objects$Culture_Group)
met_objects$Classification_Group <- as.character(met_objects$Classification_Group)
selected_variables <- met_objects %>%
  select(`Is Highlight`, Department, AccessionYear, Culture_Group, parsed_gender, parsed_year, Classifi
  mutate(Department = ifelse(is.na(Department), 'Unknown', Department)) %>%
  mutate(parsed_gender = ifelse(is.na(parsed_gender), 'Unknown', parsed_gender)) %>%
  mutate(Culture_Group = ifelse(is.na(Culture_Group), 'Unknown', Culture_Group)) %>%
  mutate(Classification_Group = ifelse(is.na(Classification_Group), 'Unknown', Classification_Group))
```

Okay now that we've got our data cleaned it's time to answer our questions.

Let's set our seed and split up our model.

```
set.seed(123)
big_met_training <- sample_n(selected_variables, nrow(selected_variables) * 0.8)
big_met_testing <- anti_join(selected_variables, big_met_training)
```

```
## Joining with `by = join_by(`Is Highlight`, Department, AccessionYear,
## Culture_Group, parsed_gender, parsed_year, Classification_Group)`
```

Setting up our baseline model...

```
big_met_testing %>%
  mutate(baseline_prediction = FALSE) %>%
  summarize(sum(baseline_prediction ==`Is Highlight`)/nrow(big_met_testing))
```

```
## # A tibble: 1 x 1
##   `sum(baseline_prediction == \`Is Highlight\`)/nrow(big_met_testing)`
##                                                                  <dbl>
## 1                                                                0.970
```

Some more data type shenanigans and making the model

```
big_met_training$Culture_Group <- as.character(big_met_training$Culture_Group)
big_met_training$Classification_Group <- as.character(big_met_training$Classification_Group)
big_met_training$Department <- as.character(big_met_training$Department)
big_met_training$parsed_gender <- as.character(big_met_training$parsed_gender)

met_model <- glm(`Is Highlight` ~ Department + AccessionYear + Culture_Group +
                      parsed_gender + parsed_year + Classification_Group,
                      data = big_met_training, family='binomial')
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

We're getting a warning here that some probabilities are numerically zero or 1. That's to be expected—some of these variables might have a near-zero relationship with the thing we're trying to measure.

Here's a summary of our model. Looks like there's quite a few significant variables!

```
summary(met_model)
```

```
##
```

```
## Call:
## glm(formula = `Is Highlight` ~ Department + AccessionYear + Culture_Group +
##     parsed_gender + parsed_year + Classification_Group, family = "binomial",
##     data = big_met_training)
##
## Coefficients:
##                                                 Estimate Std. Error
## (Intercept)                                   -2.035e+01  1.620e+00
## DepartmentArms and Armor                      -1.184e+00  2.776e-01
## DepartmentArts of Africa, Oceania, and the Americas -1.248e+00  2.287e-01
## DepartmentAsian Art                           -5.796e-01  2.015e-01
## DepartmentCostume Institute                    1.459e+01  3.001e+03
## DepartmentDrawings and Prints                 -2.427e+00  2.908e-01
## DepartmentEgyptian Art                         1.545e+01  3.001e+03
## DepartmentEuropean Paintings                   2.021e-02  2.536e-01
## DepartmentEuropean Sculpture and Decorative Arts -2.206e+00  2.370e-01
## DepartmentGreek and Roman Art                 -8.664e-01  4.800e-01
## DepartmentIslamic Art                         -5.814e-01  2.508e-01
## DepartmentMedieval Art                        -9.152e-01  2.439e-01
## DepartmentModern and Contemporary Art         -9.765e-01  2.410e-01
## DepartmentMusical Instruments                  3.033e-01  3.897e-01
## DepartmentPhotographs                         -1.071e+00  3.634e-01
## DepartmentRobert Lehman Collection             4.299e-01  2.265e-01
## DepartmentThe American Wing                    1.691e+01  3.001e+03
## DepartmentThe Cloisters                       -5.423e-01  2.510e-01
## DepartmentThe Libraries                        4.067e+01  5.348e+03
## AccessionYear                                  8.596e-03  7.937e-04
## Culture_GroupAmerican Or European             -1.288e+00  1.009e+00
## Culture_GroupBritish                           9.387e-02  2.049e-01
## Culture_GroupChina                            -1.674e+00  2.960e-01
## Culture_GroupChinese                          -1.078e+00  4.546e-01
## Culture_GroupCoptic                           -1.501e+01  4.826e+02
## Culture_GroupCypriot                          -1.206e+00  6.739e-01
## Culture_GroupDutch                             7.188e-01  5.177e-01
## Culture_GroupEtruscan                          1.468e-01  6.094e-01
## Culture_GroupEuropean                         -1.462e+01  3.390e+02
## Culture_GroupFrench                            5.191e-01  1.451e-01
## Culture_GroupGerman                            1.003e+00  1.958e-01
## Culture_GroupGreek                             9.417e-01  3.733e-01
## Culture_GroupIndian                            1.322e-01  4.742e-01
## Culture_GroupIndonesia                        -4.219e-01  4.010e-01
## Culture_GroupIran                             -6.082e-01  4.539e-01
## Culture_GroupItalian                           3.399e-01  1.985e-01
## Culture_GroupJapan                            -2.413e+00  3.940e-01
## Culture_GroupJapanese                          1.064e+00  2.508e-01
## Culture_GroupOther                             3.784e-01  1.347e-01
## Culture_GroupRoman                             4.517e-01  3.766e-01
## Culture_GroupRussian                           4.951e-01  5.959e-01
## Culture_GroupSpanish                           5.709e-01  3.317e-01
## Culture_GroupUnknown                           6.538e-02  1.489e-01
## parsed_genderMale                             -6.388e-01  1.236e-01
## parsed_genderUnknown                          -9.471e-01  9.710e-02
## parsed_year                                   -1.401e-04  7.239e-05
## Classification_GroupAlbums                    -1.416e+01  5.142e+02
```

```
## Classification_GroupArmor Parts                      -1.398e+00  1.094e+00
## Classification_GroupBooks                            -1.383e+01  1.370e+02
## Classification_GroupBronzes                           9.197e-01  6.175e-01
## Classification_GroupCeramics                          8.819e-02  4.392e-01
## Classification_GroupChordophone                       1.073e+00  2.651e-01
## Classification_GroupCodices                           1.537e+00  4.817e-01
## Classification_GroupCoins                            -1.526e+01  6.073e+02
## Classification_GroupCut Paper                         3.743e-01  1.102e+00
## Classification_GroupDrawings                          5.883e-01  4.439e-01
## Classification_GroupEnamels                           1.236e+00  5.391e-01
## Classification_GroupEphemera                         -1.406e+01  4.418e+02
## Classification_GroupGems                             -4.527e-01  7.954e-01
## Classification_GroupGlass                            -2.445e-01  4.873e-01
## Classification_GroupGold And Silver                   4.989e-01  6.572e-01
## Classification_GroupIdiophone                        -7.428e-01  4.643e-01
## Classification_GroupJade                             -1.369e+01  6.474e+02
## Classification_GroupJewelry                          -6.859e-01  5.958e-01
## Classification_GroupMedals And Plaquettes            -1.414e+01  4.369e+02
## Classification_GroupMetal                             1.259e-01  5.114e-01
## Classification_GroupMetalwork                         3.178e-01  4.422e-01
## Classification_GroupMiscellaneous                     3.508e-01  5.494e-01
## Classification_GroupNegatives                        -1.513e+01  2.505e+02
## Classification_GroupOther                             1.393e-01  4.026e-01
## Classification_GroupPaintings                         1.954e+00  4.315e-01
## Classification_GroupPhotographs                      -3.186e-01  5.051e-01
## Classification_GroupPrints                           -8.771e-01  4.806e-01
## Classification_GroupSculpture                         1.899e+00  4.338e-01
## Classification_GroupStone                             1.118e-01  4.736e-01
## Classification_GroupStone Sculpture                   2.258e+00  5.772e-01
## Classification_GroupStucco                           -1.542e+01  4.652e+02
## Classification_GroupSword Furniture                  -2.824e+00  1.106e+00
## Classification_GroupSwords                            8.169e-01  5.556e-01
## Classification_GroupTerracottas                      -7.393e-01  8.017e-01
## Classification_GroupTextiles                         -2.210e-01  4.472e-01
## Classification_GroupUnknown                          -1.603e+01  3.001e+03
## Classification_GroupVases                            -2.702e+00  6.265e-01
## Classification_GroupWood                              1.304e+00  4.708e-01
## Classification_GroupWoodwork                          1.385e+00  4.800e-01
##                                                     z value Pr(>|z|)
## (Intercept)                                        -12.561  < 2e-16 ***
## DepartmentArms and Armor                            -4.264 2.01e-05 ***
## DepartmentArts of Africa, Oceania, and the Americas -5.457 4.83e-08 ***
## DepartmentAsian Art                                 -2.876 0.004021 **
## DepartmentCostume Institute                          0.005 0.996121
## DepartmentDrawings and Prints                       -8.348  < 2e-16 ***
## DepartmentEgyptian Art                               0.005 0.995894
## DepartmentEuropean Paintings                         0.080 0.936469
## DepartmentEuropean Sculpture and Decorative Arts    -9.308  < 2e-16 ***
## DepartmentGreek and Roman Art                       -1.805 0.071110 .
## DepartmentIslamic Art                               -2.318 0.020423 *
## DepartmentMedieval Art                              -3.752 0.000175 ***
## DepartmentModern and Contemporary Art               -4.052 5.08e-05 ***
## DepartmentMusical Instruments                        0.778 0.436327
## DepartmentPhotographs                               -2.946 0.003221 **
```

```
## DepartmentRobert Lehman Collection                  1.898 0.057721 .
## DepartmentThe American Wing                         0.006 0.995504
## DepartmentThe Cloisters                            -2.161 0.030731 *
## DepartmentThe Libraries                             0.008 0.993933
## AccessionYear                                      10.831  < 2e-16 ***
## Culture_GroupAmerican Or European                  -1.276 0.201792
## Culture_GroupBritish                                0.458 0.646884
## Culture_GroupChina                                 -5.655 1.56e-08 ***
## Culture_GroupChinese                               -2.372 0.017706 *
## Culture_GroupCoptic                                -0.031 0.975192
## Culture_GroupCypriot                               -1.790 0.073486 .
## Culture_GroupDutch                                  1.388 0.164987
## Culture_GroupEtruscan                               0.241 0.809656
## Culture_GroupEuropean                              -0.043 0.965607
## Culture_GroupFrench                                 3.577 0.000347 ***
## Culture_GroupGerman                                 5.125 2.97e-07 ***
## Culture_GroupGreek                                  2.523 0.011646 *
## Culture_GroupIndian                                 0.279 0.780365
## Culture_GroupIndonesia                             -1.052 0.292710
## Culture_GroupIran                                  -1.340 0.180265
## Culture_GroupItalian                                1.712 0.086856 .
## Culture_GroupJapan                                 -6.123 9.16e-10 ***
## Culture_GroupJapanese                               4.245 2.19e-05 ***
## Culture_GroupOther                                  2.808 0.004979 **
## Culture_GroupRoman                                  1.199 0.230388
## Culture_GroupRussian                                0.831 0.406071
## Culture_GroupSpanish                                1.721 0.085206 .
## Culture_GroupUnknown                                0.439 0.660661
## parsed_genderMale                                  -5.169 2.35e-07 ***
## parsed_genderUnknown                               -9.755  < 2e-16 ***
## parsed_year                                        -1.935 0.052934 .
## Classification_GroupAlbums                         -0.028 0.978037
## Classification_GroupArmor Parts                    -1.278 0.201262
## Classification_GroupBooks                          -0.101 0.919592
## Classification_GroupBronzes                         1.489 0.136425
## Classification_GroupCeramics                        0.201 0.840861
## Classification_GroupChordophone                     4.049 5.15e-05 ***
## Classification_GroupCodices                         3.190 0.001421 **
## Classification_GroupCoins                          -0.025 0.979956
## Classification_GroupCut Paper                       0.340 0.734171
## Classification_GroupDrawings                        1.325 0.185037
## Classification_GroupEnamels                         2.293 0.021853 *
## Classification_GroupEphemera                       -0.032 0.974615
## Classification_GroupGems                           -0.569 0.569235
## Classification_GroupGlass                          -0.502 0.615782
## Classification_GroupGold And Silver                 0.759 0.447821
## Classification_GroupIdiophone                      -1.600 0.109663
## Classification_GroupJade                           -0.021 0.983131
## Classification_GroupJewelry                        -1.151 0.249687
## Classification_GroupMedals And Plaquettes          -0.032 0.974188
## Classification_GroupMetal                           0.246 0.805485
## Classification_GroupMetalwork                       0.719 0.472330
## Classification_GroupMiscellaneous                   0.639 0.523131
## Classification_GroupNegatives                      -0.060 0.951816
```

```
## Classification_GroupOther                              0.346 0.729268
## Classification_GroupPaintings                          4.529 5.93e-06 ***
## Classification_GroupPhotographs                       -0.631 0.528166
## Classification_GroupPrints                            -1.825 0.068015 .
## Classification_GroupSculpture                          4.378 1.20e-05 ***
## Classification_GroupStone                              0.236 0.813347
## Classification_GroupStone Sculpture                    3.913 9.13e-05 ***
## Classification_GroupStucco                            -0.033 0.973549
## Classification_GroupSword Furniture                   -2.553 0.010683 *
## Classification_GroupSwords                             1.470 0.141464
## Classification_GroupTerracottas                       -0.922 0.356443
## Classification_GroupTextiles                          -0.494 0.621257
## Classification_GroupUnknown                           -0.005 0.995739
## Classification_GroupVases                             -4.313 1.61e-05 ***
## Classification_GroupWood                               2.769 0.005625 **
## Classification_GroupWoodwork                           2.885 0.003919 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 22858  on 372000  degrees of freedom
## Residual deviance: 18694  on 371916  degrees of freedom
##   (15963 observations deleted due to missingness)
## AIC: 18864
##
## Number of Fisher Scoring iterations: 19
```

Our accuracy rate on the test data

```
big_met_testing %>%
  add_predictions(met_model, type='response') %>%
  mutate(predicted_outcome = ifelse(pred>=0.5, TRUE, FALSE)) %>%
  mutate(is_correct = `Is Highlight` == predicted_outcome) %>%
  summarize(mean(is_correct, na.rm=TRUE))
```

```
## # A tibble: 1 x 1
##   `mean(is_correct, na.rm = TRUE)`
##                              <dbl>
## 1                            0.973
```

So our model is slightly better than the baseline, since $0.9560013 > 0.95562$

This might seem small but over the course of our 484956 observations, the delta between those two (0.9560013-0.95562) will contribute to nearly 185 more correct predictions!

```
(0.9560013-0.95562) * 484956
```

```
## [1] 184.9137
```

Let's put our coefficients in a dataframe

```
model_summary <- summary(met_model)
coefficients_table <- model_summary$coefficients
coef_data <- as.data.frame(coefficients_table)
coef_data$Variable <- rownames(coefficients_table)
coef_data %>%
  arrange(desc(Estimate))
```

```
##                                                 Estimate   Std. Error
## DepartmentThe Libraries                     4.066600e+01 5.348208e+03
## DepartmentThe American Wing                 1.690884e+01 3.000991e+03
## DepartmentEgyptian Art                      1.544516e+01 3.000991e+03
## DepartmentCostume Institute                 1.459073e+01 3.000991e+03
## Classification_GroupStone Sculpture         2.258317e+00 5.772015e-01
## Classification_GroupPaintings               1.954255e+00 4.315272e-01
## Classification_GroupSculpture               1.899399e+00 4.338098e-01
## Classification_GroupCodices                 1.536913e+00 4.817439e-01
## Classification_GroupWoodwork                1.384702e+00 4.800268e-01
## Classification_GroupWood                    1.303711e+00 4.708479e-01
## Classification_GroupEnamels                 1.236124e+00 5.391046e-01
## Classification_GroupChordophone             1.073220e+00 2.650777e-01
## Culture_GroupJapanese                       1.064403e+00 2.507510e-01
## Culture_GroupGerman                         1.003270e+00 1.957511e-01
## Culture_GroupGreek                          9.416556e-01 3.732747e-01
## Classification_GroupBronzes                 9.196509e-01 6.175315e-01
## Classification_GroupSwords                  8.169456e-01 5.556092e-01
## Culture_GroupDutch                          7.187707e-01 5.176619e-01
## Classification_GroupDrawings                5.883166e-01 4.438770e-01
## Culture_GroupSpanish                        5.709483e-01 3.317066e-01
## Culture_GroupFrench                         5.191278e-01 1.451167e-01
## Classification_GroupGold And Silver         4.988604e-01 6.572145e-01
## Culture_GroupRussian                        4.951237e-01 5.959401e-01
## Culture_GroupRoman                          4.517349e-01 3.766468e-01
## DepartmentRobert Lehman Collection          4.298678e-01 2.265074e-01
## Culture_GroupOther                          3.783939e-01 1.347366e-01
## Classification_GroupCut Paper               3.743428e-01 1.102361e+00
## Classification_GroupMiscellaneous           3.508008e-01 5.493909e-01
## Culture_GroupItalian                        3.398936e-01 1.985104e-01
## Classification_GroupMetalwork               3.177982e-01 4.421892e-01
## DepartmentMusical Instruments               3.033293e-01 3.896776e-01
## Culture_GroupEtruscan                       1.467970e-01 6.094440e-01
## Classification_GroupOther                   1.393404e-01 4.026028e-01
## Culture_GroupIndian                         1.322351e-01 4.742260e-01
## Classification_GroupMetal                   1.259344e-01 5.114001e-01
## Classification_GroupStone                   1.118311e-01 4.736401e-01
## Culture_GroupBritish                        9.386546e-02 2.049037e-01
## Classification_GroupCeramics                8.819160e-02 4.392177e-01
## Culture_GroupUnknown                        6.537907e-02 1.489271e-01
## DepartmentEuropean Paintings                2.021299e-02 2.535844e-01
## AccessionYear                               8.595885e-03 7.936665e-04
## parsed_year                                -1.401021e-04 7.238699e-05
## Classification_GroupTextiles               -2.209708e-01 4.472469e-01
## Classification_GroupGlass                  -2.445488e-01 4.873068e-01
## Classification_GroupPhotographs            -3.186147e-01 5.050900e-01
## Culture_GroupIndonesia                     -4.219066e-01 4.009772e-01
## Classification_GroupGems                   -4.527326e-01 7.954132e-01
## DepartmentThe Cloisters                    -5.423416e-01 2.510211e-01
## DepartmentAsian Art                        -5.795502e-01 2.014789e-01
## DepartmentIslamic Art                      -5.814162e-01 2.507743e-01
## Culture_GroupIran                          -6.081938e-01 4.538961e-01
## parsed_genderMale                          -6.388359e-01 1.235890e-01
## Classification_GroupJewelry                -6.858738e-01 5.958366e-01
```

```
## Classification_GroupTerracottas                      -7.393051e-01 8.017073e-01
## Classification_GroupIdiophone                        -7.427585e-01 4.643088e-01
## DepartmentGreek and Roman Art                        -8.663718e-01 4.800441e-01
## Classification_GroupPrints                           -8.770601e-01 4.806047e-01
## DepartmentMedieval Art                               -9.151804e-01 2.439230e-01
## parsed_genderUnknown                                 -9.471442e-01 9.709542e-02
## DepartmentModern and Contemporary Art                -9.764528e-01 2.409822e-01
## DepartmentPhotographs                                -1.070541e+00 3.634131e-01
## Culture_GroupChinese                                 -1.078187e+00 4.546018e-01
## DepartmentArms and Armor                             -1.183716e+00 2.776142e-01
## Culture_GroupCypriot                                 -1.206081e+00 6.738625e-01
## DepartmentArts of Africa, Oceania, and the Americas -1.248127e+00 2.287026e-01
## Culture_GroupAmerican Or European                    -1.287765e+00 1.008854e+00
## Classification_GroupArmor Parts                      -1.397916e+00 1.093861e+00
## Culture_GroupChina                                   -1.673788e+00 2.959985e-01
## DepartmentEuropean Sculpture and Decorative Arts     -2.206081e+00 2.369996e-01
## Culture_GroupJapan                                   -2.412727e+00 3.940162e-01
## DepartmentDrawings and Prints                        -2.427079e+00 2.907509e-01
## Classification_GroupVases                            -2.702013e+00 6.264619e-01
## Classification_GroupSword Furniture                  -2.823982e+00 1.106191e+00
## Classification_GroupJade                             -1.368870e+01 6.474290e+02
## Classification_GroupBooks                            -1.383347e+01 1.370359e+02
## Classification_GroupEphemera                         -1.405919e+01 4.418168e+02
## Classification_GroupMedals And Plaquettes            -1.413798e+01 4.369465e+02
## Classification_GroupAlbums                           -1.415507e+01 5.141670e+02
## Culture_GroupEuropean                                -1.461904e+01 3.390450e+02
## Culture_GroupCoptic                                  -1.500754e+01 4.825909e+02
## Classification_GroupNegatives                        -1.513445e+01 2.504596e+02
## Classification_GroupCoins                            -1.525889e+01 6.073385e+02
## Classification_GroupStucco                           -1.542411e+01 4.651727e+02
## Classification_GroupUnknown                          -1.602714e+01 3.000991e+03
## (Intercept)                                          -2.035361e+01 1.620386e+00
##                                                            z value     Pr(>|z|)
## DepartmentThe Libraries                                0.007603669 9.939332e-01
## DepartmentThe American Wing                            0.005634420 9.955044e-01
## DepartmentEgyptian Art                                 0.005146687 9.958936e-01
## DepartmentCostume Institute                            0.004861971 9.961207e-01
## Classification_GroupStone Sculpture                    3.912527226 9.133524e-05
## Classification_GroupPaintings                          4.528696048 5.934880e-06
## Classification_GroupSculpture                          4.378415460 1.195453e-05
## Classification_GroupCodices                            3.190310677 1.421199e-03
## Classification_GroupWoodwork                           2.884633898 3.918691e-03
## Classification_GroupWood                               2.768857293 5.625327e-03
## Classification_GroupEnamels                            2.292921252 2.185254e-02
## Classification_GroupChordophone                        4.048699435 5.150304e-05
## Culture_GroupJapanese                                  4.244861395 2.187286e-05
## Culture_GroupGerman                                    5.125231076 2.971731e-07
## Culture_GroupGreek                                     2.522688288 1.164616e-02
## Classification_GroupBronzes                            1.489237070 1.364250e-01
## Classification_GroupSwords                             1.470360200 1.414642e-01
## Culture_GroupDutch                                     1.388494531 1.649865e-01
## Classification_GroupDrawings                           1.325404713 1.850370e-01
## Culture_GroupSpanish                                   1.721245154 8.520635e-02
## Culture_GroupFrench                                    3.577313777 3.471434e-04
```

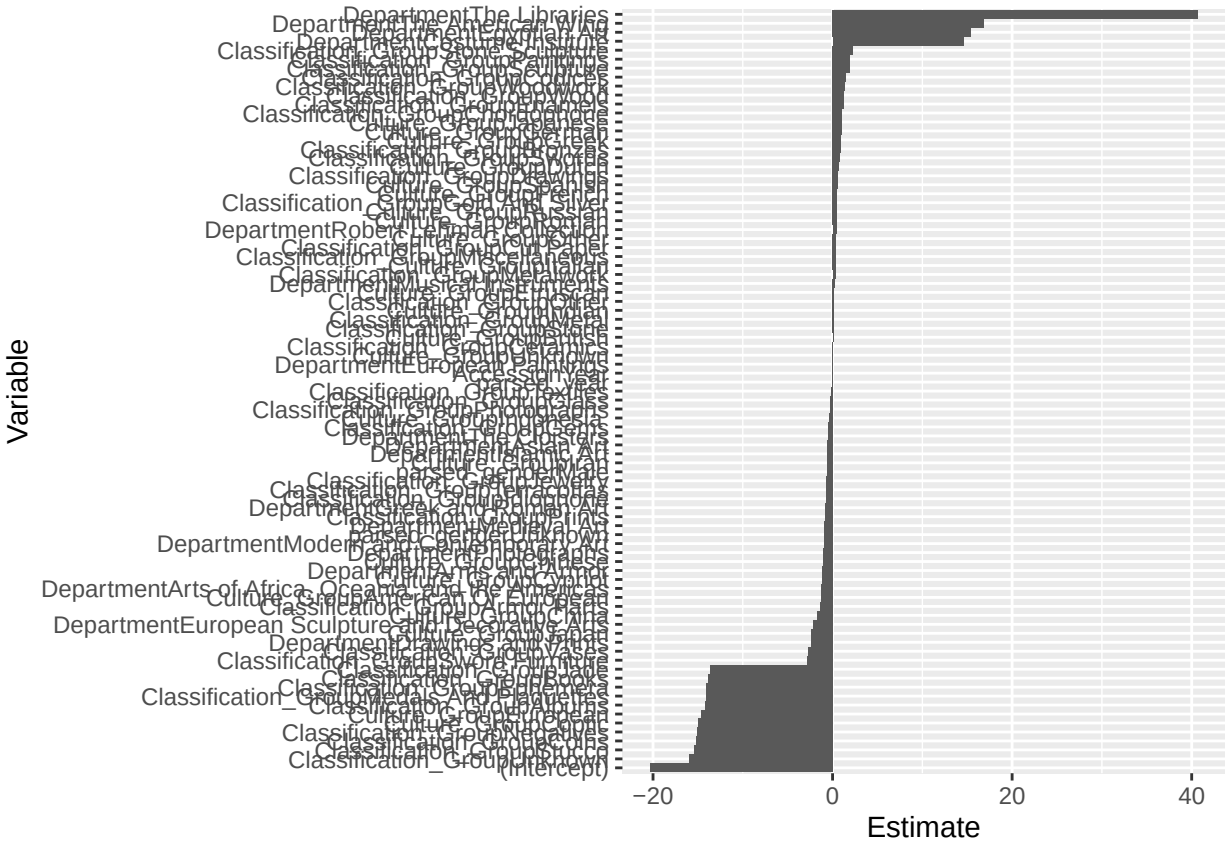
```
## Classification_GroupGold And Silver                    0.759052669 4.478211e-01
## Culture_GroupRussian                                   0.830827905 4.060709e-01
## Culture_GroupRoman                                     1.199359231 2.303883e-01
## DepartmentRobert Lehman Collection                     1.897808829 5.772127e-02
## Culture_GroupOther                                     2.808397459 4.978874e-03
## Classification_GroupCut Paper                          0.339582923 7.341706e-01
## Classification_GroupMiscellaneous                      0.638526678 5.231309e-01
## Culture_GroupItalian                                   1.712220939 8.685596e-02
## Classification_GroupMetalwork                          0.718692811 4.723302e-01
## DepartmentMusical Instruments                          0.778410804 4.363269e-01
## Culture_GroupEtruscan                                  0.240870431 8.096555e-01
## Classification_GroupOther                              0.346099023 7.292683e-01
## Culture_GroupIndian                                    0.278844055 7.803645e-01
## Classification_GroupMetal                              0.246254211 8.054855e-01
## Classification_GroupStone                              0.236109921 8.133474e-01
## Culture_GroupBritish                                   0.458095476 6.468838e-01
## Classification_GroupCeramics                           0.200792471 8.408609e-01
## Culture_GroupUnknown                                   0.439000421 6.606612e-01
## DepartmentEuropean Paintings                           0.079709135 9.364686e-01
## AccessionYear                                         10.830600551 2.465331e-27
## parsed_year                                           -1.935459880 5.293389e-02
## Classification_GroupTextiles                          -0.494069007 6.212574e-01
## Classification_GroupGlass                             -0.501837413 6.157819e-01
## Classification_GroupPhotographs                       -0.630807756 5.281662e-01
## Culture_GroupIndonesia                                -1.052195991 2.927096e-01
## Classification_GroupGems                              -0.569179174 5.692346e-01
## DepartmentThe Cloisters                               -2.160541556 3.073077e-02
## DepartmentAsian Art                                   -2.876481477 4.021358e-03
## DepartmentIslamic Art                                 -2.318483529 2.042306e-02
## Culture_GroupIran                                     -1.339940768 1.802646e-01
## parsed_genderMale                                     -5.169036988 2.353033e-07
## Classification_GroupJewelry                           -1.151110416 2.496868e-01
## Classification_GroupTerracottas                       -0.922163342 3.564434e-01
## Classification_GroupIdiophone                         -1.599708017 1.096634e-01
## DepartmentGreek and Roman Art                         -1.804775473 7.110982e-02
## Classification_GroupPrints                            -1.824909432 6.801470e-02
## DepartmentMedieval Art                                -3.751922850 1.754835e-04
## parsed_genderUnknown                                  -9.754777927 1.759854e-22
## DepartmentModern and Contemporary Art                 -4.051971230 5.078791e-05
## DepartmentPhotographs                                 -2.945797826 3.221229e-03
## Culture_GroupChinese                                  -2.371716286 1.770568e-02
## DepartmentArms and Armor                              -4.263887635 2.009005e-05
## Culture_GroupCypriot                                  -1.789802560 7.348566e-02
## DepartmentArts of Africa, Oceania, and the Americas   -5.457421557 4.830985e-08
## Culture_GroupAmerican Or European                     -1.276463483 2.017917e-01
## Classification_GroupArmor Parts                       -1.277965306 2.012617e-01
## Culture_GroupChina                                    -5.654719365 1.561011e-08
## DepartmentEuropean Sculpture and Decorative Arts      -9.308376126 1.298016e-20
## Culture_GroupJapan                                    -6.123420017 9.158784e-10
## DepartmentDrawings and Prints                         -8.347622636 6.965069e-17
## Classification_GroupVases                             -4.313132447 1.609576e-05
## Classification_GroupSword Furniture                   -2.552887268 1.068341e-02
## Classification_GroupJade                              -0.021143170 9.831314e-01
## Classification_GroupBooks                             -0.100947776 9.195919e-01
```

```
## Classification_GroupEphemera                    -0.031821306 9.746146e-01
## Classification_GroupMedals And Plaquettes        -0.032356309 9.741879e-01
## Classification_GroupAlbums                       -0.027530096 9.780369e-01
## Culture_GroupEuropean                            -0.043118283 9.656072e-01
## Culture_GroupCoptic                              -0.031097840 9.751915e-01
## Classification_GroupNegatives                    -0.060426704 9.518158e-01
## Classification_GroupCoins                        -0.025124189 9.799559e-01
## Classification_GroupStucco                       -0.033157810 9.735487e-01
## Classification_GroupUnknown                      -0.005340617 9.957388e-01
## (Intercept)                                    -12.560963289 3.460933e-36
##                                                                          Variab
## DepartmentThe Libraries                                          DepartmentThe Librari
## DepartmentThe American Wing                                    DepartmentThe American Wi
## DepartmentEgyptian Art                                           DepartmentEgyptian A
## DepartmentCostume Institute                                    DepartmentCostume Institu
## Classification_GroupStone Sculpture                  Classification_GroupStone Sculptu
## Classification_GroupPaintings                          Classification_GroupPaintin
## Classification_GroupSculpture                          Classification_GroupSculptu
## Classification_GroupCodices                              Classification_GroupCodic
## Classification_GroupWoodwork                           Classification_GroupWoodwo
## Classification_GroupWood                                   Classification_GroupWo
## Classification_GroupEnamels                            Classification_GroupEnamo
## Classification_GroupChordophone                      Classification_GroupChordopho
## Culture_GroupJapanese                                        Culture_GroupJapano
## Culture_GroupGerman                                            Culture_GroupGerm
## Culture_GroupGreek                                             Culture_GroupGre
## Classification_GroupBronzes                            Classification_GroupBronz
## Classification_GroupSwords                             Classification_GroupSwor
## Culture_GroupDutch                                             Culture_GroupDut
## Classification_GroupDrawings                           Classification_GroupDrawin
## Culture_GroupSpanish                                           Culture_GroupSpani
## Culture_GroupFrench                                            Culture_GroupFren
## Classification_GroupGold And Silver                  Classification_GroupGold And Silv
## Culture_GroupRussian                                           Culture_GroupRussi
## Culture_GroupRoman                                             Culture_GroupRom
## DepartmentRobert Lehman Collection                   DepartmentRobert Lehman Collecti
## Culture_GroupOther                                             Culture_GroupOth
## Classification_GroupCut Paper                          Classification_GroupCut Pap
## Classification_GroupMiscellaneous                    Classification_GroupMiscellaneo
## Culture_GroupItalian                                           Culture_GroupItali
## Classification_GroupMetalwork                          Classification_GroupMetalwo
## DepartmentMusical Instruments                          DepartmentMusical Instrumen
## Culture_GroupEtruscan                                        Culture_GroupEtrusc
## Classification_GroupOther                                Classification_GroupOth
## Culture_GroupIndian                                            Culture_GroupIndi
## Classification_GroupMetal                                Classification_GroupMet
## Classification_GroupStone                                Classification_GroupSto
## Culture_GroupBritish                                           Culture_GroupBriti
## Classification_GroupCeramics                           Classification_GroupCerami
## Culture_GroupUnknown                                           Culture_GroupUnkno
## DepartmentEuropean Paintings                           DepartmentEuropean Paintin
## AccessionYear                                                          AccessionYe
## parsed_year                                                              parsed_ye
## Classification_GroupTextiles                           Classification_GroupTextil
```

```
## Classification_GroupGlass                                                  Classification_GroupGl
## Classification_GroupPhotographs                                        Classification_GroupPhotograp
## Culture_GroupIndonesia                                                      Culture_GroupIndones
## Classification_GroupGems                                                   Classification_GroupG
## DepartmentThe Cloisters                                                    DepartmentThe Cloist
## DepartmentAsian Art                                                            DepartmentAsian
## DepartmentIslamic Art                                                        DepartmentIslamic
## Culture_GroupIran                                                             Culture_GroupI
## parsed_genderMale                                                             parsed_genderM
## Classification_GroupJewelry                                              Classification_GroupJewe
## Classification_GroupTerracottas                                        Classification_GroupTerracot
## Classification_GroupIdiophone                                            Classification_GroupIdioph
## DepartmentGreek and Roman Art                                            DepartmentGreek and Roman
## Classification_GroupPrints                                                 Classification_GroupPri
## DepartmentMedieval Art                                                         DepartmentMedieval
## parsed_genderUnknown                                                         parsed_genderUnkn
## DepartmentModern and Contemporary Art                            DepartmentModern and Contemporary
## DepartmentPhotographs                                                       DepartmentPhotograp
## Culture_GroupChinese                                                        Culture_GroupChin
## DepartmentArms and Armor                                                  DepartmentArms and Ar
## Culture_GroupCypriot                                                        Culture_GroupCypr
## DepartmentArts of Africa, Oceania, and the Americas DepartmentArts of Africa, Oceania, and the Americ
## Culture_GroupAmerican Or European                                  Culture_GroupAmerican Or Europ
## Classification_GroupArmor Parts                                      Classification_GroupArmor Pa
## Culture_GroupChina                                                            Culture_GroupCh
## DepartmentEuropean Sculpture and Decorative Arts      DepartmentEuropean Sculpture and Decorative A
## Culture_GroupJapan                                                          Culture_GroupJap
## DepartmentDrawings and Prints                                          DepartmentDrawings and Pri
## Classification_GroupVases                                                   Classification_GroupVa
## Classification_GroupSword Furniture                                Classification_GroupSword Furnitu
## Classification_GroupJade                                                     Classification_GroupJ
## Classification_GroupBooks                                                   Classification_GroupBo
## Classification_GroupEphemera                                             Classification_GroupEphem
## Classification_GroupMedals And Plaquettes                      Classification_GroupMedals And Plaquet
## Classification_GroupAlbums                                                  Classification_GroupAlb
## Culture_GroupEuropean                                                       Culture_GroupEurop
## Culture_GroupCoptic                                                           Culture_GroupCop
## Classification_GroupNegatives                                          Classification_GroupNegati
## Classification_GroupCoins                                                   Classification_GroupCo
## Classification_GroupStucco                                                Classification_GroupStu
## Classification_GroupUnknown                                              Classification_GroupUnkn
## (Intercept)                                                                                (Interce
```

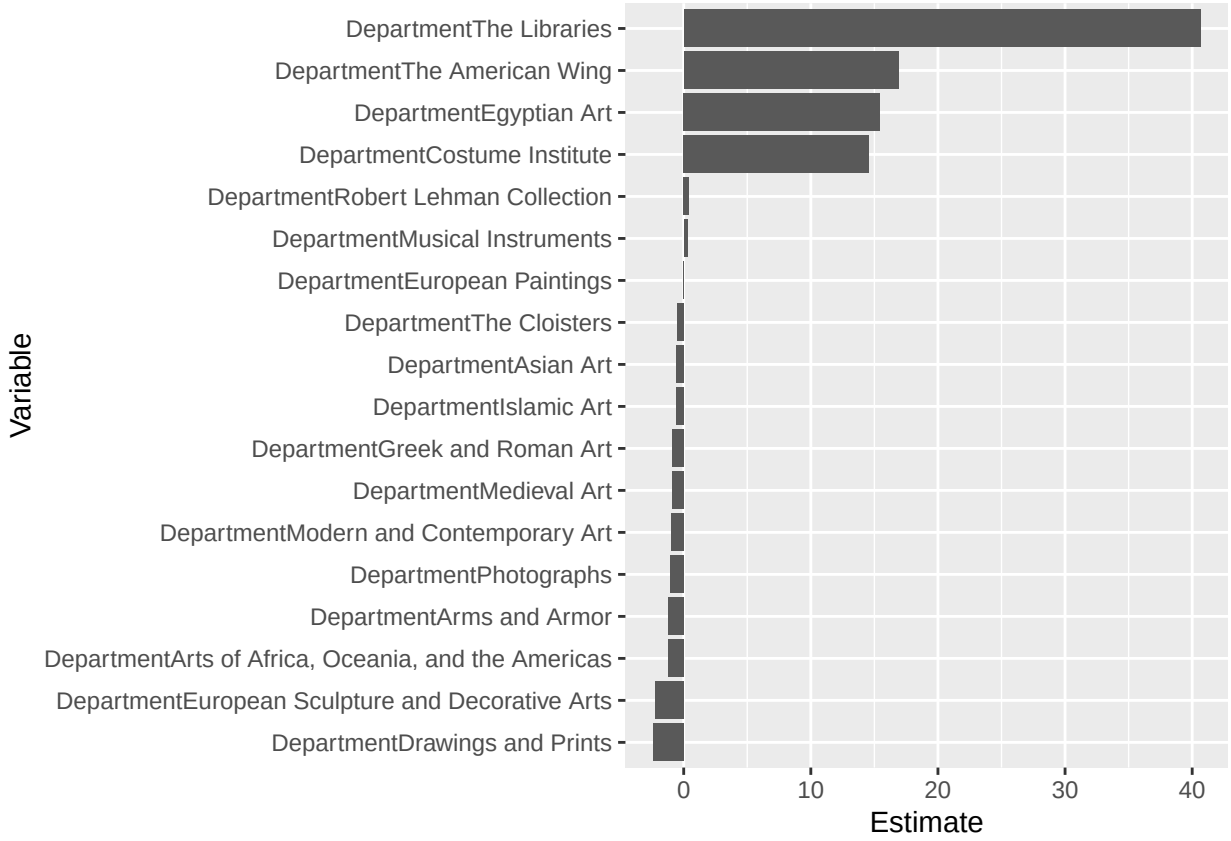
Now let's graph them!

```
coef_data %>%
  mutate(Variable = reorder(Variable, Estimate)) %>%
  ggplot() +
  geom_bar(aes(y = Variable, x = Estimate), stat = "identity")
```

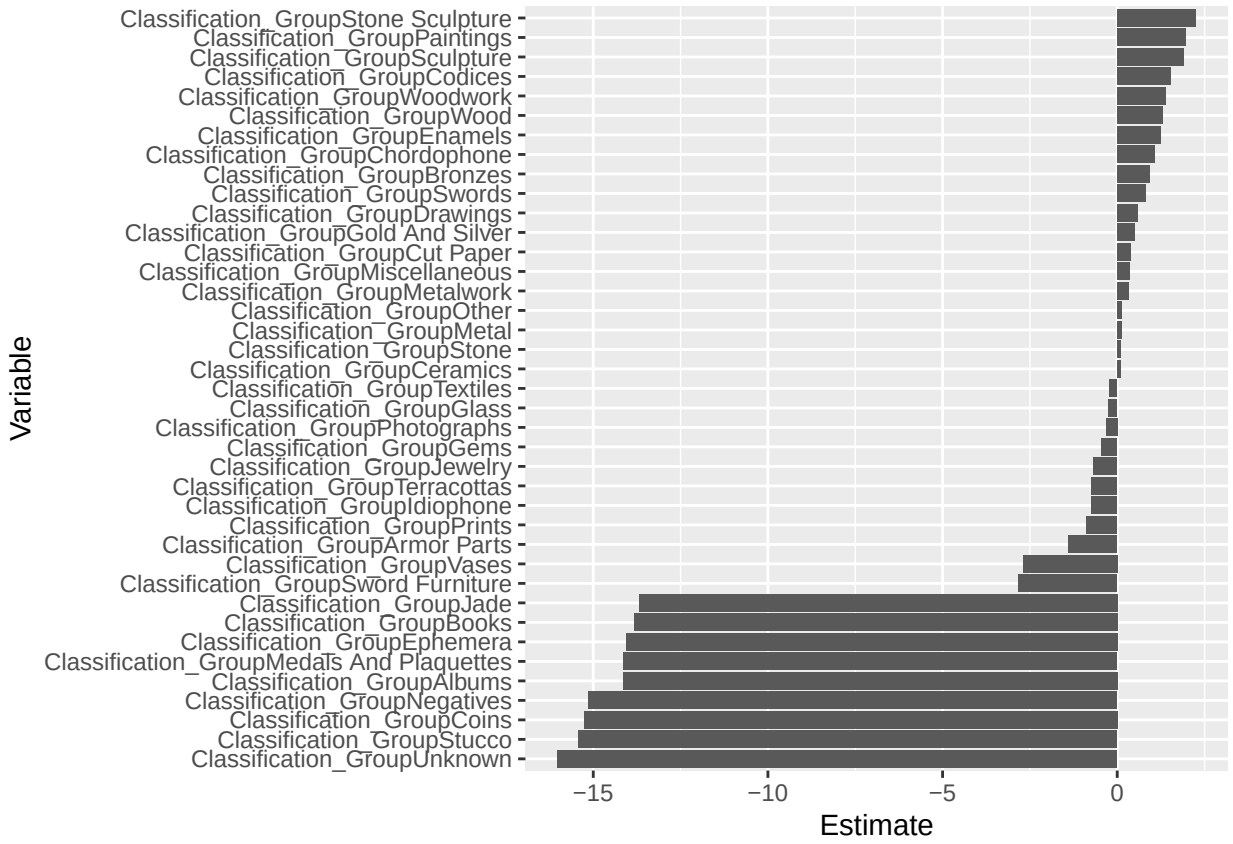


Let's make plots for each of the types of categorical variables so we can see them more clearly.

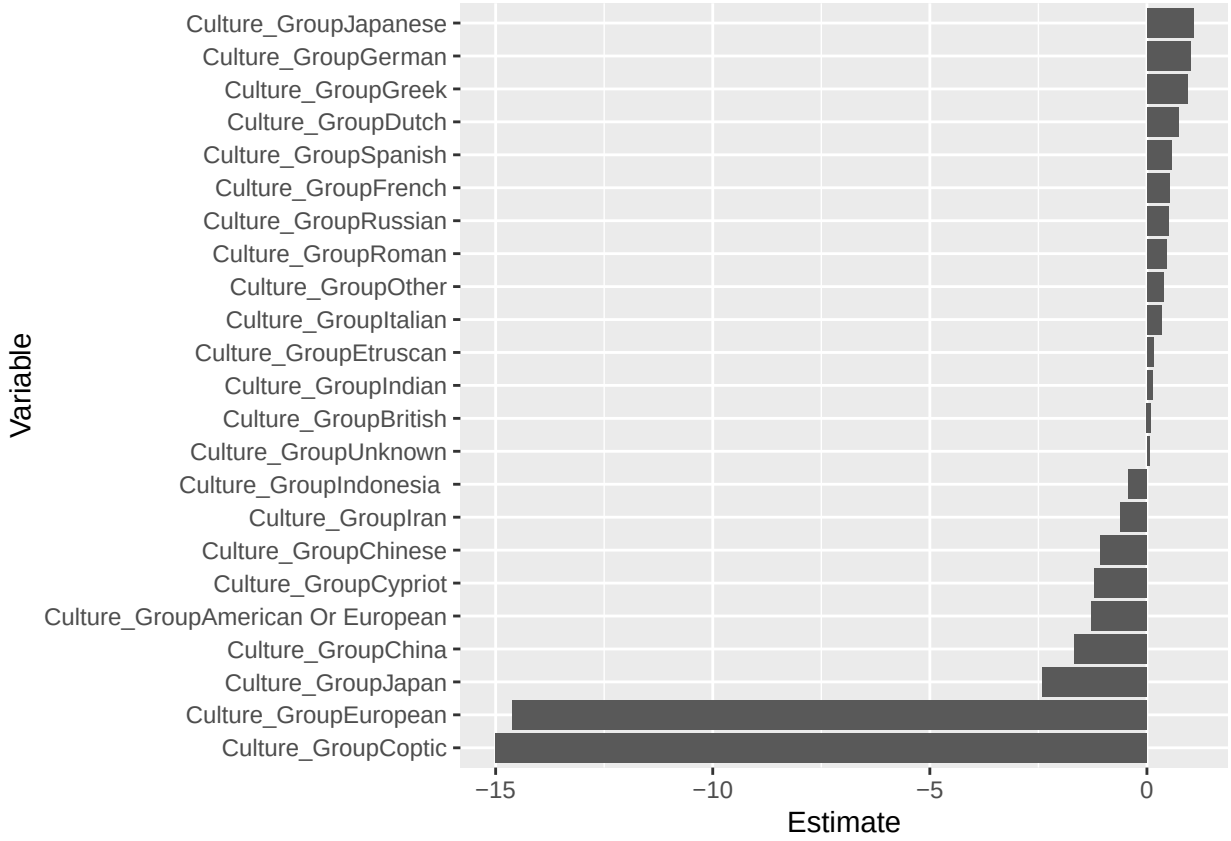
```
coef_data %>%
  filter(str_detect(Variable, "Department")) %>%
  mutate(Variable = reorder(Variable, Estimate)) %>%
  ggplot() +
  geom_bar(aes(y = Variable, x = Estimate), stat = "identity")
```




```
coef_data %>%
  filter(str_detect(Variable, "Classification")) %>%
  mutate(Variable = reorder(Variable, Estimate)) %>%
  ggplot() +
  geom_bar(aes(y = Variable, x = Estimate), stat = "identity")
```



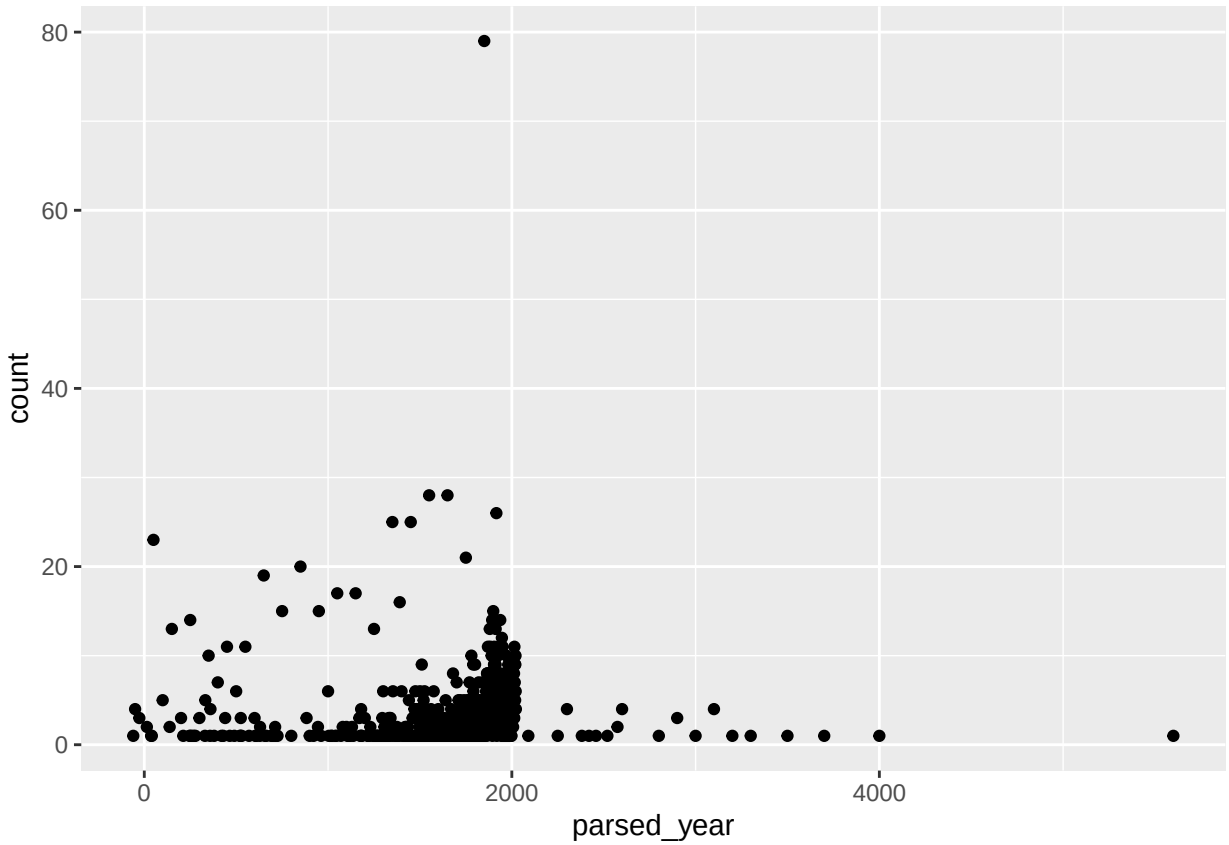

```
coef_data %>%
  filter(str_detect(Variable, "Culture")) %>%
  mutate(Variable = reorder(Variable, Estimate)) %>%
  ggplot() +
  geom_bar(aes(y = Variable, x = Estimate), stat = "identity")
```



Let's do some more analysis on time period though, because even if there's no linear relationship there could still be a relationship.

```
big_met_training %>%
  filter(`Is Highlight` == TRUE) %>%
  select(parsed_year, `Is Highlight`) %>%
  group_by(parsed_year) %>%
  summarize(count = n()) %>%
  ggplot() +
  geom_point(aes(x = parsed_year, y = count))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

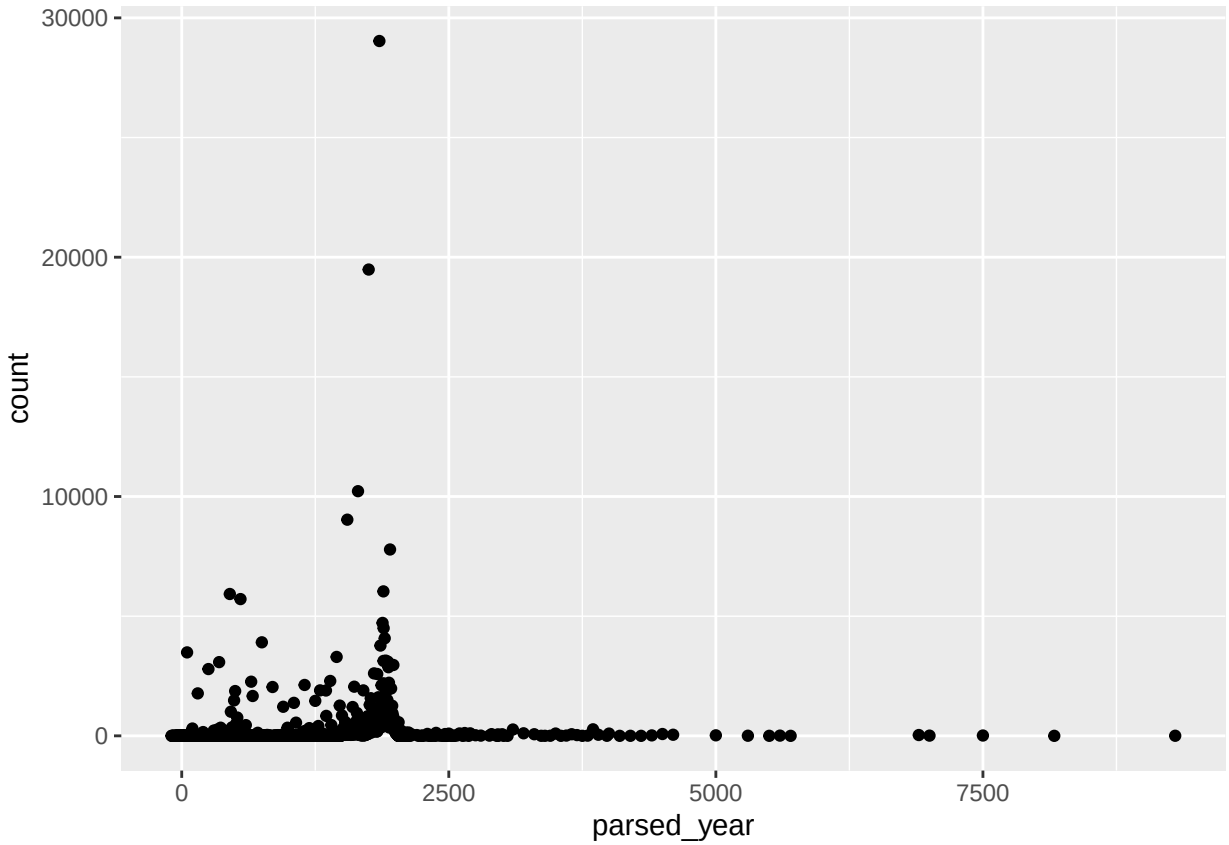


We get a warning on this and the next few graphs that one row was removed. Don't see any need to worry about that since we have hundreds of thousands of rows.

That looks encouraging, but let's make sure we're not just looking at a trend in the overall number of objects.

```
big_met_training %>%
  group_by(parsed_year) %>%
  summarize(count = n()) %>%
  ggplot() +
  geom_point(aes(x=parsed_year, y=count))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



These plots look really weird but they're just artifacts of how we've parsed the year data. This is roughly how we should expect it to look—let's check out the distinct values.

```
big_met_training %>%
  select(parsed_year) %>%
  distinct() %>%
  arrange(desc(parsed_year))
```

```
## # A tibble: 1,496 x 1
##    parsed_year
##          <dbl>
##  1        9300
##  2        8168
##  3        7500
##  4        7000
##  5        6900
##  6        5700
##  7        5600
##  8        5500
##  9        5300
## 10        5000
## # i 1,486 more rows
```

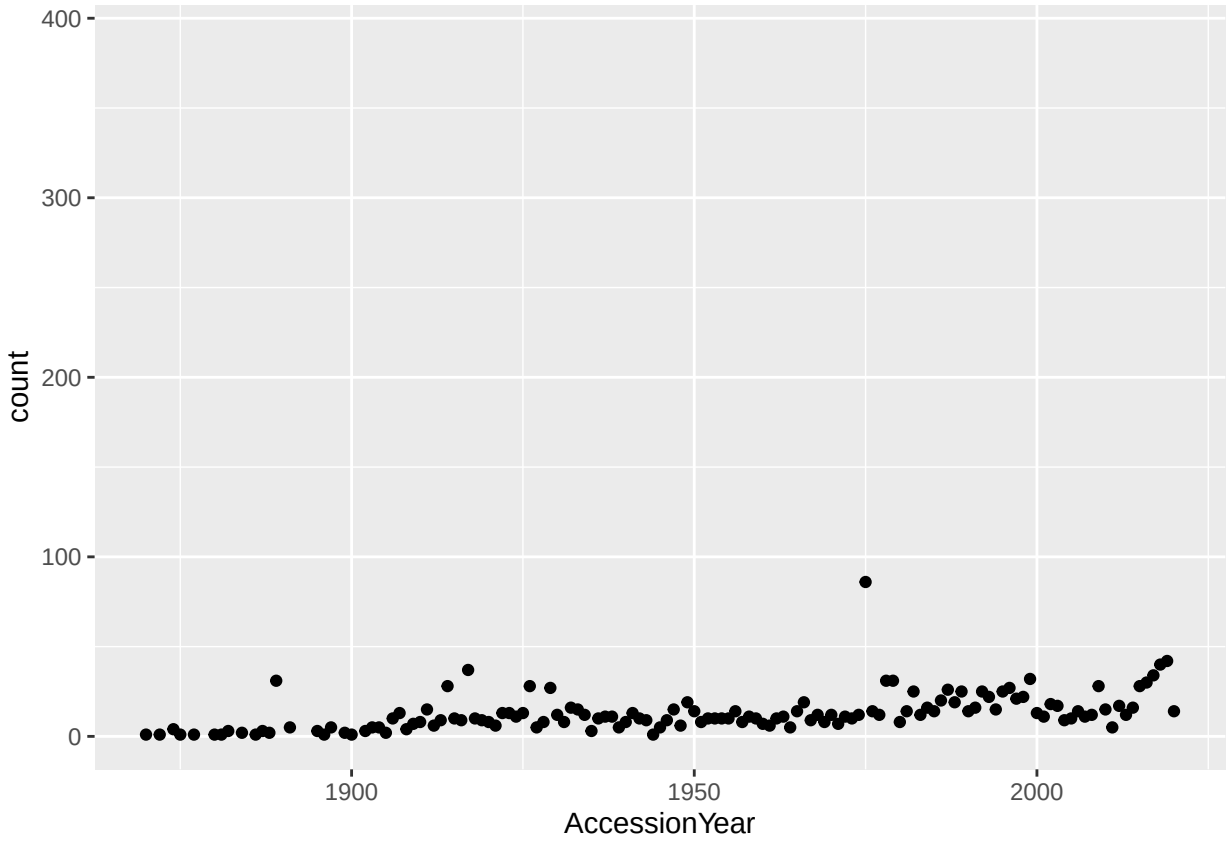
So it seems like there isn't any special relationship with time.

Let's check it out with our AccessionYear too.

```
big_met_training %>%
  filter(`Is Highlight` == TRUE) %>%
```

```
  select(AccessionYear, `Is Highlight`) %>%
  group_by(AccessionYear) %>%
  summarize(count = n()) %>%
  ggplot() +
  geom_point(aes(x = AccessionYear, y = count))
```

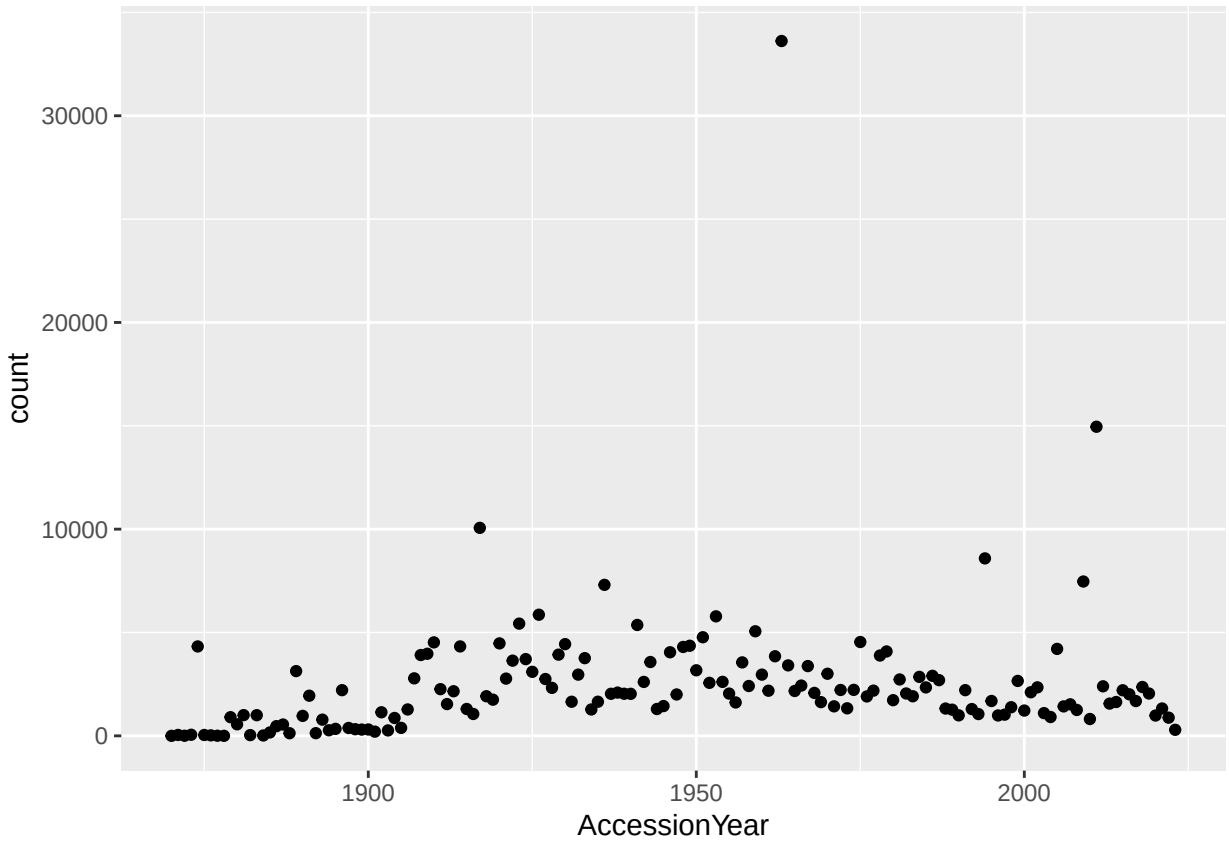
```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



This looks like a linear relationship but let's check the overall data.

```
big_met_training %>%
  group_by(AccessionYear) %>%
  summarize(count = n()) %>%
  ggplot() +
  geom_point(aes(x=AccessionYear, y=count))
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



No huge differences here but it's somewhat interesting that all the outliers disappeared.

Conclusions

Main Questions

Question 1: Which variables have the highest correlation with `isHighlight`? Let's see which of our variables are significant.

```
coef_data$Significant <- coef_data$`Pr(>|z|)` < 0.05
significant_vars <- coef_data[coef_data$Significant, ]
significant_vars %>%
  arrange(desc(Estimate))
```

```
##                                                 Estimate   Std. Error
## Classification_GroupStone Sculpture          2.258316653 0.5772015178
## Classification_GroupPaintings                1.954255435 0.4315271801
## Classification_GroupSculpture                1.899399405 0.4338097704
## Classification_GroupCodices                  1.536912709 0.4817439006
## Classification_GroupWoodwork                 1.384701586 0.4800268023
## Classification_GroupWood                     1.303710548 0.4708478663
## Classification_GroupEnamels                  1.236124422 0.5391046121
## Classification_GroupChordophone              1.073219967 0.2650777080
## Culture_GroupJapanese                        1.064403208 0.2507509927
## Culture_GroupGerman                          1.003269862 0.1957511470
## Culture_GroupGreek                           0.941655592 0.3732746515
## Culture_GroupFrench                          0.519127829 0.1451166605
## Culture_GroupOther                           0.378393914 0.1347365962
## AccessionYear                                0.008595885 0.0007936665
```

```
## DepartmentThe Cloisters                                -0.542341575 0.2510211263
## DepartmentAsian Art                                   -0.579550189 0.2014788531
## DepartmentIslamic Art                                 -0.581416181 0.2507743419
## parsed_genderMale                                     -0.638835890 0.1235889571
## DepartmentMedieval Art                                -0.915180366 0.2439230235
## parsed_genderUnknown                                  -0.947144231 0.0970954170
## DepartmentModern and Contemporary Art                 -0.976452802 0.2409821656
## DepartmentPhotographs                                 -1.070541480 0.3634130864
## Culture_GroupChinese                                  -1.078186555 0.4546018262
## DepartmentArms and Armor                              -1.183715599 0.2776141635
## DepartmentArts of Africa, Oceania, and the Americas  -1.248126701 0.2287026370
## Culture_GroupChina                                    -1.673788227 0.2959984605
## DepartmentEuropean Sculpture and Decorative Arts     -2.206081038 0.2369995592
## Culture_GroupJapan                                    -2.412726975 0.3940162472
## DepartmentDrawings and Prints                         -2.427078710 0.2907508899
## Classification_GroupVases                             -2.702013331 0.6264619425
## Classification_GroupSword Furniture                   -2.823981773 1.1061913345
## (Intercept)                                          -20.353605233 1.6203856953
##                                                          z value     Pr(>|z|)
## Classification_GroupStone Sculpture                     3.912527 9.133524e-05
## Classification_GroupPaintings                           4.528696 5.934880e-06
## Classification_GroupSculpture                           4.378415 1.195453e-05
## Classification_GroupCodices                             3.190311 1.421199e-03
## Classification_GroupWoodwork                            2.884634 3.918691e-03
## Classification_GroupWood                                2.768857 5.625327e-03
## Classification_GroupEnamels                             2.292921 2.185254e-02
## Classification_GroupChordophone                         4.048699 5.150304e-05
## Culture_GroupJapanese                                   4.244861 2.187286e-05
## Culture_GroupGerman                                     5.125231 2.971731e-07
## Culture_GroupGreek                                      2.522688 1.164616e-02
## Culture_GroupFrench                                     3.577314 3.471434e-04
## Culture_GroupOther                                      2.808397 4.978874e-03
## AccessionYear                                          10.830601 2.465331e-27
## DepartmentThe Cloisters                                -2.160542 3.073077e-02
## DepartmentAsian Art                                    -2.876481 4.021358e-03
## DepartmentIslamic Art                                  -2.318484 2.042306e-02
## parsed_genderMale                                      -5.169037 2.353033e-07
## DepartmentMedieval Art                                 -3.751923 1.754835e-04
## parsed_genderUnknown                                   -9.754778 1.759854e-22
## DepartmentModern and Contemporary Art                  -4.051971 5.078791e-05
## DepartmentPhotographs                                  -2.945798 3.221229e-03
## Culture_GroupChinese                                   -2.371716 1.770568e-02
## DepartmentArms and Armor                               -4.263888 2.009005e-05
## DepartmentArts of Africa, Oceania, and the Americas   -5.457422 4.830985e-08
## Culture_GroupChina                                     -5.654719 1.561011e-08
## DepartmentEuropean Sculpture and Decorative Arts      -9.308376 1.298016e-20
## Culture_GroupJapan                                     -6.123420 9.158784e-10
## DepartmentDrawings and Prints                          -8.347623 6.965069e-17
## Classification_GroupVases                              -4.313132 1.609576e-05
## Classification_GroupSword Furniture                    -2.552887 1.068341e-02
## (Intercept)                                           -12.560963 3.460933e-36
##                                                                                Variab
## Classification_GroupStone Sculpture                                 Classification_GroupStone Sculptu
## Classification_GroupPaintings                                             Classification_GroupPaintin
```

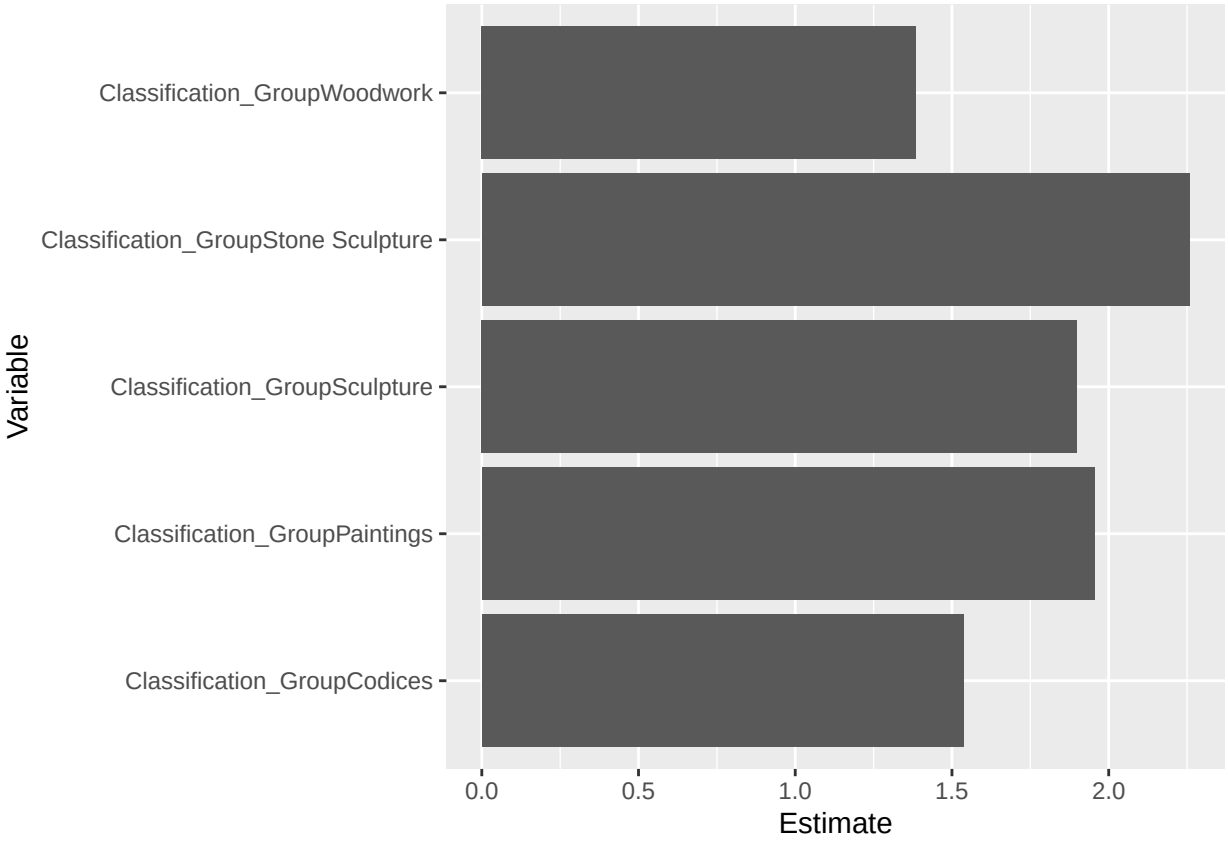
```
## Classification_GroupSculpture                                                  Classification_GroupSculptu
## Classification_GroupCodices                                                      Classification_GroupCodi
## Classification_GroupWoodwork                                                    Classification_GroupWoodwo
## Classification_GroupWood                                                           Classification_GroupWo
## Classification_GroupEnamels                                                     Classification_GroupEnamo
## Classification_GroupChordophone                                              Classification_GroupChordopho
## Culture_GroupJapanese                                                              Culture_GroupJapano
## Culture_GroupGerman                                                                Culture_GroupGerm
## Culture_GroupGreek                                                                 Culture_GroupGre
## Culture_GroupFrench                                                                Culture_GroupFren
## Culture_GroupOther                                                                 Culture_GroupOth
## AccessionYear                                                                           AccessionYe
## DepartmentThe Cloisters                                                         DepartmentThe Cloiste
## DepartmentAsian Art                                                                DepartmentAsian A
## DepartmentIslamic Art                                                            DepartmentIslamic A
## parsed_genderMale                                                                    parsed_genderMa
## DepartmentMedieval Art                                                           DepartmentMedieval A
## parsed_genderUnknown                                                               parsed_genderUnkno
## DepartmentModern and Contemporary Art                                 DepartmentModern and Contemporary A
## DepartmentPhotographs                                                            DepartmentPhotograp
## Culture_GroupChinese                                                               Culture_GroupChino
## DepartmentArms and Armor                                                        DepartmentArms and Arm
## DepartmentArts of Africa, Oceania, and the Americas DepartmentArts of Africa, Oceania, and the Americ
## Culture_GroupChina                                                                   Culture_GroupChi
## DepartmentEuropean Sculpture and Decorative Arts       DepartmentEuropean Sculpture and Decorative Ar
## Culture_GroupJapan                                                                 Culture_GroupJap
## DepartmentDrawings and Prints                                                 DepartmentDrawings and Prin
## Classification_GroupVases                                                        Classification_GroupVas
## Classification_GroupSword Furniture                                    Classification_GroupSword Furnitu
## (Intercept)                                                                                 (Intercep
##                                                    Significant
## Classification_GroupStone Sculpture                       TRUE
## Classification_GroupPaintings                             TRUE
## Classification_GroupSculpture                             TRUE
## Classification_GroupCodices                               TRUE
## Classification_GroupWoodwork                              TRUE
## Classification_GroupWood                                  TRUE
## Classification_GroupEnamels                               TRUE
## Classification_GroupChordophone                           TRUE
## Culture_GroupJapanese                                     TRUE
## Culture_GroupGerman                                       TRUE
## Culture_GroupGreek                                        TRUE
## Culture_GroupFrench                                       TRUE
## Culture_GroupOther                                        TRUE
## AccessionYear                                             TRUE
## DepartmentThe Cloisters                                   TRUE
## DepartmentAsian Art                                       TRUE
## DepartmentIslamic Art                                     TRUE
## parsed_genderMale                                         TRUE
## DepartmentMedieval Art                                    TRUE
## parsed_genderUnknown                                      TRUE
## DepartmentModern and Contemporary Art                     TRUE
## DepartmentPhotographs                                     TRUE
## Culture_GroupChinese                                      TRUE
```

```
## DepartmentArms and Armor                                    TRUE
## DepartmentArts of Africa, Oceania, and the Americas         TRUE
## Culture_GroupChina                                          TRUE
## DepartmentEuropean Sculpture and Decorative Arts            TRUE
## Culture_GroupJapan                                          TRUE
## DepartmentDrawings and Prints                               TRUE
## Classification_GroupVases                                   TRUE
## Classification_GroupSword Furniture                         TRUE
## (Intercept)                                                 TRUE
```

We have 32 significant variables—all of which have z scores above 2.5. We can see that the top five most positively associated variables are all Classification Groups— Stone Sculpture, Paintings, Sculpture, Codices, and Woodwork.

Let's graph them!

```
significant_vars %>%
  filter(Estimate>1.35) %>%
  ggplot() + geom_bar(aes(y=Variable, x=Estimate), stat='identity')
```



Question 2: If I'm artist and want the MET to think my art is "a popular and important artwork in the collection" what kind of art should I make? Let's remember- the variables we analyzed were `Department`, `AccessionYear`, `Culture`, `Artist End Date`, `Artist Gender`, `Object Date`, and `Classification`.

Some of these variables are beyond your control—for instance when you're born, what your gender is, and when you make your art. Culture might be under your control somewhat if you're a multicultural artist, but for most people you probably can't manipulate that variable either. So all we're down to is Classification and Department.

Let's look at the significant variables in those dimensions.

```
significant_vars %>%
  filter(str_detect(Variable, "Department|Classification")) %>%
  arrange(desc(Estimate))
```

```
##                                                    Estimate Std. Error
## Classification_GroupStone Sculpture                2.2583167  0.5772015
## Classification_GroupPaintings                      1.9542554  0.4315272
## Classification_GroupSculpture                      1.8993994  0.4338098
## Classification_GroupCodices                        1.5369127  0.4817439
## Classification_GroupWoodwork                       1.3847016  0.4800268
## Classification_GroupWood                           1.3037105  0.4708479
## Classification_GroupEnamels                        1.2361244  0.5391046
## Classification_GroupChordophone                    1.0732200  0.2650777
## DepartmentThe Cloisters                           -0.5423416  0.2510211
## DepartmentAsian Art                               -0.5795502  0.2014789
## DepartmentIslamic Art                             -0.5814162  0.2507743
## DepartmentMedieval Art                            -0.9151804  0.2439230
## DepartmentModern and Contemporary Art             -0.9764528  0.2409822
## DepartmentPhotographs                             -1.0705415  0.3634131
## DepartmentArms and Armor                          -1.1837156  0.2776142
## DepartmentArts of Africa, Oceania, and the Americas -1.2481267  0.2287026
## DepartmentEuropean Sculpture and Decorative Arts  -2.2060810  0.2369996
## DepartmentDrawings and Prints                     -2.4270787  0.2907509
## Classification_GroupVases                         -2.7020133  0.6264619
## Classification_GroupSword Furniture               -2.8239818  1.1061913
##                                                     z value     Pr(>|z|)
## Classification_GroupStone Sculpture                3.912527 9.133524e-05
## Classification_GroupPaintings                      4.528696 5.934880e-06
## Classification_GroupSculpture                      4.378415 1.195453e-05
## Classification_GroupCodices                        3.190311 1.421199e-03
## Classification_GroupWoodwork                       2.884634 3.918691e-03
## Classification_GroupWood                           2.768857 5.625327e-03
## Classification_GroupEnamels                        2.292921 2.185254e-02
## Classification_GroupChordophone                    4.048699 5.150304e-05
## DepartmentThe Cloisters                           -2.160542 3.073077e-02
## DepartmentAsian Art                               -2.876481 4.021358e-03
## DepartmentIslamic Art                             -2.318484 2.042306e-02
## DepartmentMedieval Art                            -3.751923 1.754835e-04
## DepartmentModern and Contemporary Art             -4.051971 5.078791e-05
## DepartmentPhotographs                             -2.945798 3.221229e-03
## DepartmentArms and Armor                          -4.263888 2.009005e-05
## DepartmentArts of Africa, Oceania, and the Americas -5.457422 4.830985e-08
## DepartmentEuropean Sculpture and Decorative Arts  -9.308376 1.298016e-20
## DepartmentDrawings and Prints                     -8.347623 6.965069e-17
## Classification_GroupVases                         -4.313132 1.609576e-05
## Classification_GroupSword Furniture               -2.552887 1.068341e-02
##                                                                                Variab
## Classification_GroupStone Sculpture                                 Classification_GroupStone Sculptu
## Classification_GroupPaintings                                             Classification_GroupPaintin
## Classification_GroupSculpture                                             Classification_GroupSculptu
## Classification_GroupCodices                                                 Classification_GroupCodi
## Classification_GroupWoodwork                                               Classification_GroupWoodwo
## Classification_GroupWood                                                       Classification_GroupWo
```

```
## Classification_GroupEnamels                                                 Classification_GroupEnam
## Classification_GroupChordophone                                          Classification_GroupChordoph
## DepartmentThe Cloisters                                                        DepartmentThe Cloist
## DepartmentAsian Art                                                               DepartmentAsian A
## DepartmentIslamic Art                                                           DepartmentIslamic A
## DepartmentMedieval Art                                                         DepartmentMedieval A
## DepartmentModern and Contemporary Art                                DepartmentModern and Contemporary A
## DepartmentPhotographs                                                         DepartmentPhotograp
## DepartmentArms and Armor                                                    DepartmentArms and Arm
## DepartmentArts of Africa, Oceania, and the Americas DepartmentArts of Africa, Oceania, and the Americ
## DepartmentEuropean Sculpture and Decorative Arts       DepartmentEuropean Sculpture and Decorative Ar
## DepartmentDrawings and Prints                                        DepartmentDrawings and Prin
## Classification_GroupVases                                                  Classification_GroupVas
## Classification_GroupSword Furniture                                 Classification_GroupSword Furnitu
##                                                      Significant
## Classification_GroupStone Sculpture                         TRUE
## Classification_GroupPaintings                               TRUE
## Classification_GroupSculpture                               TRUE
## Classification_GroupCodices                                 TRUE
## Classification_GroupWoodwork                                TRUE
## Classification_GroupWood                                    TRUE
## Classification_GroupEnamels                                 TRUE
## Classification_GroupChordophone                             TRUE
## DepartmentThe Cloisters                                     TRUE
## DepartmentAsian Art                                         TRUE
## DepartmentIslamic Art                                       TRUE
## DepartmentMedieval Art                                      TRUE
## DepartmentModern and Contemporary Art                       TRUE
## DepartmentPhotographs                                       TRUE
## DepartmentArms and Armor                                    TRUE
## DepartmentArts of Africa, Oceania, and the Americas         TRUE
## DepartmentEuropean Sculpture and Decorative Arts            TRUE
## DepartmentDrawings and Prints                               TRUE
## Classification_GroupVases                                   TRUE
## Classification_GroupSword Furniture                         TRUE
```

It seems like most of our variables are actually part of these two categories.

Looking at the most positively and negatively associated variables, it seems like you're best off if you make Stone Sculptures or Paintings, but manage to avoid being put in the "Drawing and Prints" or "European Sculpture and Decorative Arts" departments.

Question 3: What time periods and cultures are the best-represented among the Met's "popular and important artwork"? Let's check out our significant variables again!

```
significant_vars %>%
  filter(str_detect(Variable, "Culture|Year")) %>%
  arrange(desc(Estimate))
```

```
##                          Estimate   Std. Error   z value     Pr(>|z|)
## Culture_GroupJapanese  1.064403208 0.2507509927  4.244861 2.187286e-05
## Culture_GroupGerman    1.003269862 0.1957511470  5.125231 2.971731e-07
## Culture_GroupGreek     0.941655592 0.3732746515  2.522688 1.164616e-02
## Culture_GroupFrench    0.519127829 0.1451166605  3.577314 3.471434e-04
## Culture_GroupOther     0.378393914 0.1347365962  2.808397 4.978874e-03
## AccessionYear          0.008595885 0.0007936665 10.830601 2.465331e-27
```

```
## Culture_GroupChinese  -1.078186555 0.4546018262 -2.371716 1.770568e-02
## Culture_GroupChina    -1.673788227 0.2959984605 -5.654719 1.561011e-08
## Culture_GroupJapan    -2.412726975 0.3940162472 -6.123420 9.158784e-10
##                                    Variable Significant
## Culture_GroupJapanese Culture_GroupJapanese        TRUE
## Culture_GroupGerman     Culture_GroupGerman        TRUE
## Culture_GroupGreek       Culture_GroupGreek        TRUE
## Culture_GroupFrench     Culture_GroupFrench        TRUE
## Culture_GroupOther       Culture_GroupOther        TRUE
## AccessionYear                 AccessionYear        TRUE
## Culture_GroupChinese   Culture_GroupChinese        TRUE
## Culture_GroupChina       Culture_GroupChina        TRUE
## Culture_GroupJapan       Culture_GroupJapan        TRUE
```

Something weird is going on here—the Japanese culture group is doing great, but the Japan culture group is doing really poorly. You'd expect them to not be significant if this were the case but they both are—in fact they both have Z scores over 4! So there must be some quirk in whether the MET labels art as Japanese or Japan that is really indicative of the propensity of that art to be a highlight.

The other best-off culture groups are the Germans, the Greeks, and the French.

Interestingly, while AccessionYear has a low coefficient, it's extremely robust, with P functionally equivalent to zero. This is a little surprising because parsed_year is not even a significant variable. The difference between the two is that parsed_year describes when the art was made, whereas AccessionYear describes the year the MET acquired the art. So you're better off having the MET acquire your art later on, although not much.

Next Steps

An interesting next step would be to try this sort of analysis on other large art museums and see if the model holds. I looked at a few other museums but none I could find had an equivalent to the `Is Highlight` variable.

Another place you could go is start tracking what the museums put in their gallery's versus what they put in the warehouses and see how that correlates to the sort of variables I looked at—that would get at a similar question but probably in a more universally applicable way.

Sources

Code: https://github.com/RowanGray472/MET-data-analysis Data: https://github.com/metmuseum/openaccess