# Replication: Semantics derived automatically from language corpora necessarily contain human biases

*Alisabeth Marsteller*

## Dependencies

```
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)

source('./R/weat.R')
source('./R/wefat.R')
```

## Data

```
words_df <- fread("./data/glove.840B.300d.txt")
```

## Baseline: Replication of Associations That Are Universally Accepted

### Flowers & Insects

```
#TARGETS
flowers <- c("aster", "clover", "hyacinth", "marigold", "poppy", "azalea", "crocus",
             "iris", "orchid", "rose", "bluebell", "daffodil", "lilac", "pansy",
             "tulip", "buttercup", "daisy", "lily", "peony", "violet", "carnation",
             "magnolia", "zinnia")

insects <- c("ant", "caterpillar", "flea", "locust", "spider", "bedbug", "centipede",
             "fly", "maggot", "tarantula", "bee", "cockroach", "gnat", "mosquito",
             "termite", "beetle", "cricket", "hornet", "moth", "wasp", "blackfly",
             "dragonfly", "horsefly", "roach", "weevil")

#ATTRIBUTES
pleasant <- c("caress", "freedom", "health", "love", "peace", "cheer", "friend", "heaven",
              "loyal", "pleasure", "diamond", "gentle", "honest", "lucky", "rainbow",
              "diploma", "gift", "honor", "miracle", "sunrise", "family", "happy",
              "laughter", "paradise", "vacation")

unpleasant <- c("abuse", "crash", "filth", "murder", "sickness", "accident", "death",
                "grief", "poison", "stink", "assault", "disaster", "hatred", "pollute",
                "tragedy", "divorce", "jail", "poverty", "ugly", "cancer", "kill",
                "rotten", "vomit", "agony", "prison")
#p-value, effect size
y <- weat(words_df, flowers, insects, pleasant, unpleasant)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 4.91740438047333e-08"
```

```
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.54387416816756"
```

### Musical Instruments & Weapons

```
#TARGETS
music <- c("bagpipe", "cello", "guitar", "lute", "trombone", "banjo", "clarinet",
           "harmonica", "mandolin", "trumpet", "bassoon", "drum", "harp", "oboe",
           "tuba", "bell", "fiddle", "harpsichord", "piano", "viola", "bongo",
           "flute", "horn", "saxophone", "violin")

weapon <- c("arrow", "club", "gun", "missile", "spear", "axe", "dagger", "harpoon",
            "pistol", "sword", "blade", "dynamite", "hatchet", "rifle", "tank", "bomb",
            "firearm", "knife", "shotgun", "teargas", "cannon", "grenade", "mace",
            "slingshot", "whip")

#ATTRIBUTES
pleasant <- c("caress", "freedom", "health", "love", "peace", "cheer", "friend", "heaven",
              "loyal", "pleasure", "diamond", "gentle", "honest", "lucky", "rainbow",
              "diploma", "gift", "honor", "miracle", "sunrise", "family", "happy",
              "laughter", "paradise", "vacation")

unpleasant <- c("abuse", "crash", "filth", "murder", "sickness", "accident", "death",
                "grief", "poison", "stink", "assault", "disaster", "hatred", "pollute",
                "tragedy", "divorce", "jail", "poverty", "ugly", "cancer", "kill",
                "rotten", "vomit", "agony", "prison")
#p-value, effect size
y <- weat(words_df, music, weapon, pleasant, unpleasant)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```r
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 2.09551784585302e-09"
```

```r
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.53398870351849"
```

# Racial Biases

### Replicating Implicit Associations for Valence

```r
#TARGETS
euro <- c("Adam", "Harry", "Josh", "Roger", "Alan", "Frank", "Justin", "Ryan", "Andrew",
          "Jack", "Matthew", "Stephen", "Brad", "Greg", "Paul", "Jonathan", "Peter",
          "Amanda", "Courtney", "Heather", "Melanie", "Katie", "Betsy", "Kristin",
          "Nancy", "Stephanie", "Ellen", "Lauren", "Colleen", "Emily", "Megan", "Rachel")

afr <- c("Alonzo", "Jamel", "Theo", "Alphonse", "Jerome", "Leroy", "Torrance", "Darnell",
         "Lamar", "Lionel", "Tvree", "Deion", "Lamont", "Malik", "Terrence", "Tyrone",
         "Lavon", "Marcellus", "Wardell", "Nichelle", "Shereen", "Ebony", "Latisha",
         "Shaniqua", "Jasmine", "Tanisha", "Tia", "Lakisha", "Latoya", "Yolanda",
         "Malika", "Yvette")
#ATTRIBUTES
pleasant <- c("caress", "freedom", "health", "love", "peace", "cheer", "friend", "heaven",
              "loyal", "pleasure", "diamond", "gentle", "honest", "lucky", "rainbow",
              "diploma", "gift", "honor", "miracle", "sunrise", "family", "happy",
              "laughter", "paradise", "vacation")

unpleasant <- c("abuse", "crash", "filth", "murder", "sickness", "accident", "death",
                "grief", "poison", "stink", "assault", "disaster", "hatred", "pollute",
                "tragedy", "bomb", "divorce", "jail", "poverty", "ugly", "cancer",
                "evil", "kill", "rotten", "vomit")
#p-value, effect size
y <- weat(words_df, euro, afr, pleasant, unpleasant)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```r
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 1.3632089440908e-06"
```

```r
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.39181358492378"
```

### Replicating the Bertrand and Mullainathan (2004) Résumé Study

```r
#TARGETS
euro <- c("Brad", "Brendan", "Geoffrey", "Greg", "Brett", "Matthew", "Neil", "Todd",
          "Allison", "Anne", "Carrie", "Emily", "Jill", "Laurie", "Meredith", "Sarah")
```

```r
afr <- c("Darnell", "Hakim", "Jermaine", "Kareem", "Jamal", "Leroy", "Rasheed", "Tyrone",
         "Aisha", "Ebony", "Keisha", "Kenya", "Lakisha", "Latoya", "Tamika", "Tanisha")
#ATTRIBUTES
pleasant <- c("caress", "freedom", "health", "love", "peace", "cheer", "friend", "heaven",
              "loyal", "pleasure", "diamond", "gentle", "honest", "lucky", "rainbow",
              "diploma", "gift", "honor", "miracle", "sunrise", "family", "happy",
              "laughter", "paradise", "vacation")

unpleasant <- c("abuse", "crash", "filth", "murder", "sickness", "accident", "death",
                "grief", "poison", "stink", "assault", "disaster", "hatred", "pollute",
                "tragedy", "bomb", "divorce", "jail", "poverty", "ugly", "cancer",
                "evil", "kill", "rotten", "vomit")

# Updated Nosek et al. Attributes
pleasantness <- c("joy", "love", "peace", "wonderful", "pleasure", "friend", "laughter", "happy")

unpleasantness <- c("agony", "terrible", "horrible", "nasty", "evil", "war", "awful", "failure")

#p-value, effect size
y <- weat(words_df, euro, afr, pleasant, unpleasant)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```r
paste0("Greenwald et al. Attributes:")
```

```
## [1] "Greenwald et al. Attributes:"
```

```r
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 2.11767659864568e-07"
```

```r
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.48050571481421"
```

```r
#p-value, effect size
y <- weat(words_df, euro, afr, pleasantness, unpleasantness)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```r
paste0("Nosek et al. Attributes:")
```

```
## [1] "Nosek et al. Attributes:"
```

```r
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 0.000154180886198602"
```

```r
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.25799162696255"
```

# Gender Biases

### Replicating Implicit Associations for Career and Family

```
#TARGETS
male <- c("John", "Paul", "Mike", "Kevin", "Steve", "Greg", "Jeff", "Bill")
fem <- c("Amy", "Joan", "Lisa", "Sarah", "Diana", "Kate", "Ann", "Donna")
#ATTRIBUTES
car <- c("executive", "management", "professional", "corporation", "salary", "office",
         "business", "career")
fam <- c("home", "parents", "children", "family", "cousins", "marriage", "wedding",
         "relatives")

#p-value, effect size
y <- weat(words_df, male, fem, car, fam)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 3.88510884955811e-06"
```

```
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.81391462368104"
```

### Replicating Implicit Associations for Arts and Mathematics

```
#TARGETS
math <- c("math", "algebra", "geometry", "calculus", "equations", "computation",
          "numbers", "addition")
art <- c("poetry", "art", "dance", "literature", "novel", "symphony", "drama",
         "sculpture")
#ATTRIBUTES
male <- c("male", "man", "boy", "brother", "he", "him", "his", "son")
fem <- c("female", "woman", "girl", "sister", "she", "her", "hers", "daughter")

#p-value, effect size
y <- weat(words_df, math, art, male, fem)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 0.0203409153207412"
```

```
paste0("effect size: ", y[2])
```

```
## [1] "effect size: 1.05501478731626"
```

### Replicating Implicit Associations for Arts and Sciences

```
#TARGETS
science <- c("science", "technology", "physics", "chemistry", "Einstein", "NASA",
             "experiment", "astronomy")
art <- c("poetry", "art", "Shakespeare", "dance", "literature", "novel", "symphony",
         "drama")
#ATTRIBUTES
male <- c("brother", "father", "uncle", "grandfather", "son", "he", "his", "him")
female <- c("sister", "mother", "aunt", "grandmother", "daughter", "she", "hers", "her")

#p-value, effect size
y <- weat(words_df, science, art, male, female)
```

```
## [1] "Computing test statistic and effect size..."
## [1] "Computing p-value..."
```

```
paste0("p-value: ", y[1])
```

```
## [1] "p-value: 0.000384884933432325"
```

```
paste0("effect size: ", y[2])
```

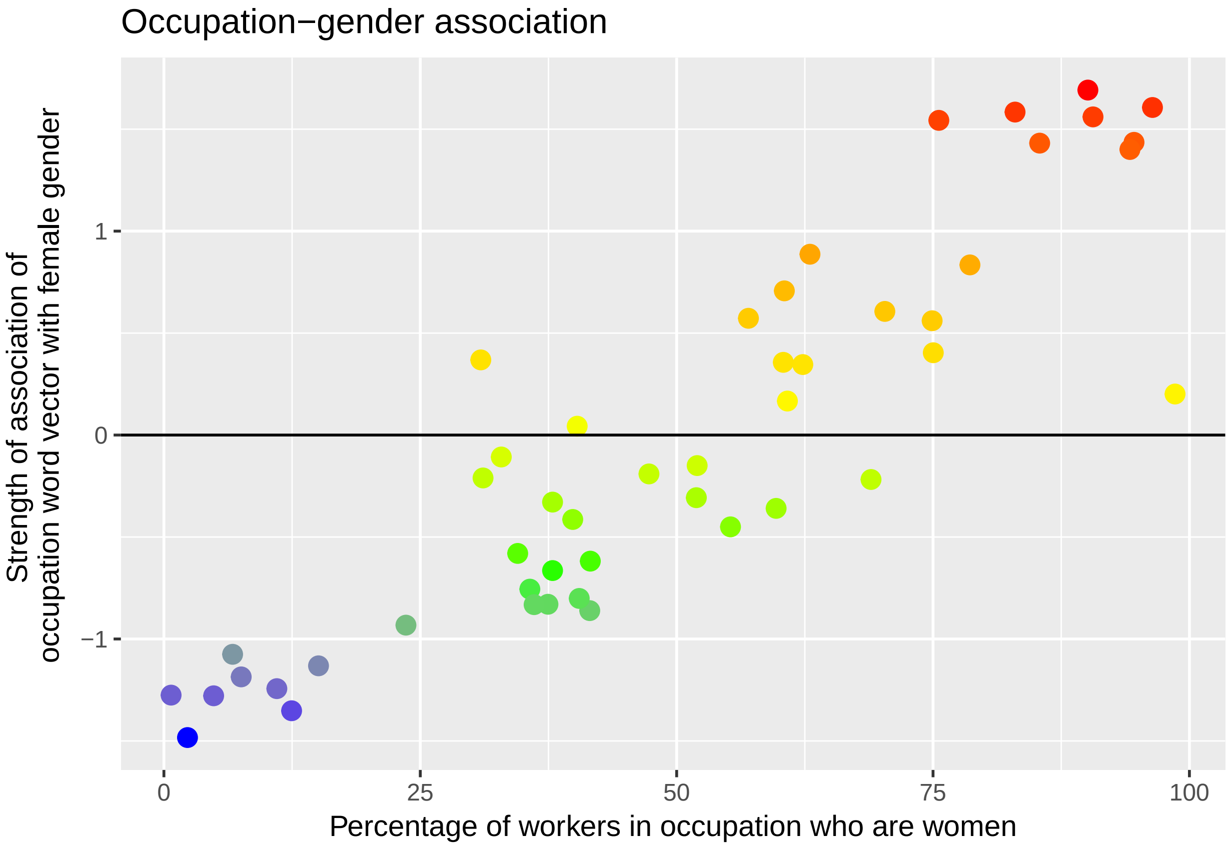
```
## [1] "effect size: 1.2374533926249"
```

### Comparison to Real-World Data: Occupational Statistics

```
#TARGET
occup <- c("technician", "accountant", "supervisor", "engineer", "worker", "educator",
           "clerk", "counselor", "inspector", "mechanic", "manager", "therapist",
           "administrator", "salesperson", "receptionist", "librarian", "advisor",
           "pharmacist", "janitor", "psychologist", "physician", "carpenter", "nurse",
           "investigator", "bartender", "specialist", "electrician", "officer",
           "pathologist", "teacher", "lawyer", "planner", "practitioner", "plumber",
           "instructor", "surgeon", "veterinarian", "paramedic", "examiner", "chemist",
           "machinist", "appraiser", "nutritionist", "architect", "hairdresser", "baker",
           "programmer", "paralegal", "hygienist", "scientist")
#ATTRIBUTES
female <- c("female", "woman", "girl", "sister", "she", "her", "hers", "daughter")

male <- c("male", "man", "boy", "brother", "he", "him", "his", "son")

assoc_fem_occ <- wefat(words_df, occup, female, male)

#retrieve Labor Statistics Data 2015
stats <- read.csv("./data/2015census.csv")
#keep only occupation/percent women and delete rows with no data
stats <- stats %>%
  select(occupation, Women) %>%
  mutate_all(funs(as.character(.))) %>%
  filter(Women != "-")

#get the mean percent of women in each target occupation word from census data
perc <- c()
```

```r
for (i in 1:length(occup)) {
  careers <- stats[grep(occup[i], stats[,1]), ]
  career2 <- as.numeric(careers[,2])
  percwomen <- mean(career2)
  perc[i] <- percwomen
}

#create data frame with word associations and percentage female workers in each occupation
fem_occup_tot <- as.data.frame(cbind(perc, as.data.frame(assoc_fem_occ)))

#plot strength of association vs. percent female workers in each occupation
ggplot(data = fem_occup_tot, aes(x = perc, y = assoc_fem_occ, color = assoc_fem_occ)) +
  scale_colour_gradientn(colors = c('blue', 'green','yellow', 'orange','red')) +
  geom_point(size = 3) +
  xlab("Percentage of workers in occupation who are women") +
  ylab("Strength of association of
       occupation word vector with female gender") +
  ggtitle("Occupation-gender association") +
  geom_hline(yintercept = 0) +
  theme(legend.position = "none")
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



```r
ggsave("./results/fig/occupational_statistics.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

```
#Pearson correlation coefficient
p_fem_occ <- cor(fem_occup_tot[,1], fem_occup_tot[,2], use = "pairwise.complete.obs",
                 method = "pearson")
p_fem_occ
```

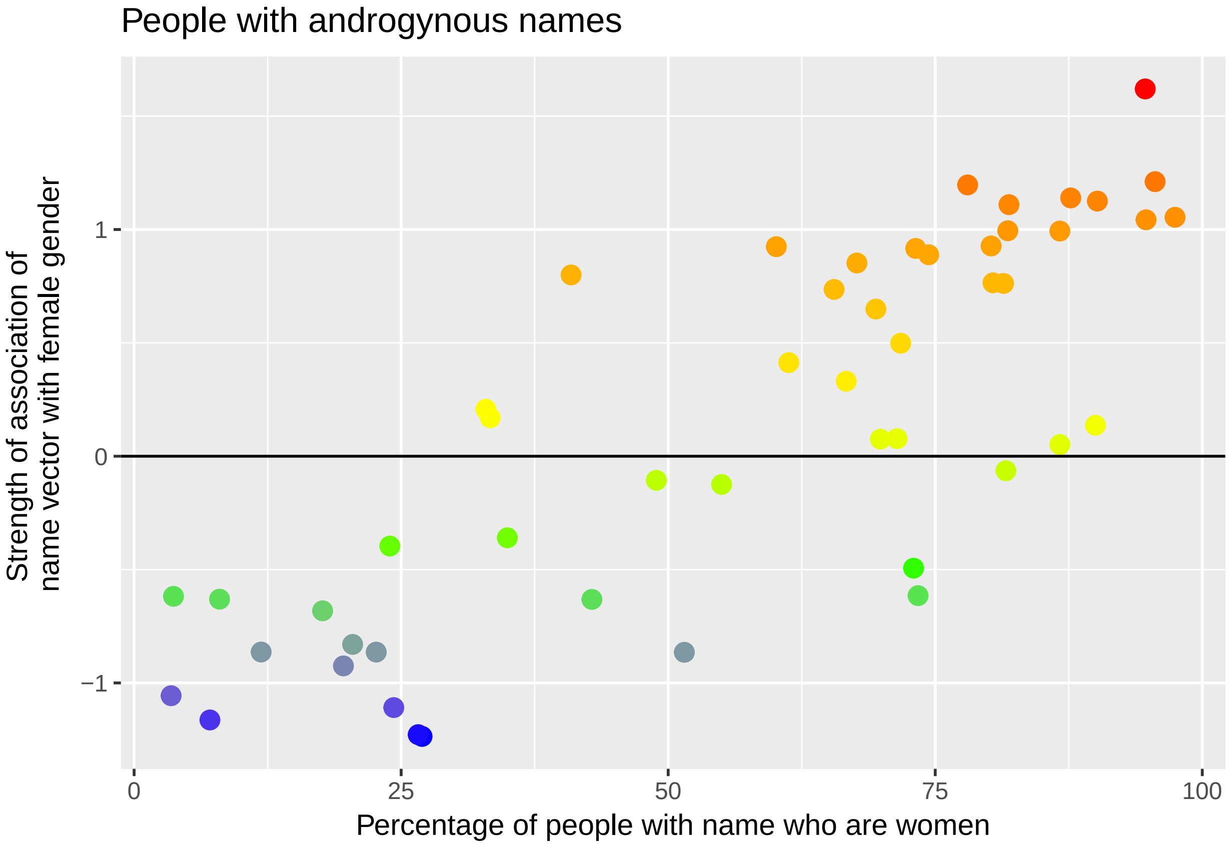
```
## [1] 0.8953161
```

### Comparison to Real-World Data: Androgynous Names

```
#TARGET
names <- c("Kelly", "Tracy", "Jamie", "Jackie", "Jesse", "Courtney", "Lynn", "Taylor",
           "Leslie", "Shannon", "Stacey", "Jessie", "Shawn", "Stacy", "Casey", "Bobby",
           "Terry", "Lee", "Ashley", "Eddie", "Chris", "Jody", "Pat", "Carey", "Willie",
           "Morgan", "Robbie", "Joan", "Alexis", "Kris", "Frankie", "Bobbie", "Dale",
           "Robin", "Billie", "Adrian", "Kim", "Jaime", "Jean", "Francis", "Marion",
           "Dana", "Rene", "Johnnie", "Jordan", "Carmen", "Ollie", "Dominique", "Jimmie",
           "Shelby")
#ATTRIBUTES
#male/female attributes (same as above)
female <- c("female", "woman", "girl", "sister", "she", "her", "hers", "daughter")

male <- c("male", "man", "boy", "brother", "he", "him", "his", "son")

#get association strengths between target words and both attributes
assoc_fem_names <- wefat(words_df, names, female, male)
```

```
#Analysis of results
#1990 stats: percentage of women with certain name
stats_femnames <- read.table("./data/female.first1990.txt",
                             header = TRUE) %>%
  select(Name, freq, rank) %>%
  mutate(Name = as.character(Name))

#1990 stats: percentage of men with certain name
stats_malenames <- read.table("./data/male.first1990.txt",
                              header = TRUE) %>%
  select(Name, freq, rank)


#get from data the percent of women with certain names
freq_femnames <- c()
for (i in 1:length(names)) {
  freq <- stats_femnames[grep(names[i], stats_femnames[,1]), 'freq']
  freq_femnames[i] <- freq*100
}

#get from data the percent of men with certain names
freq_malenames <- c()
for (i in 1:length(names)) {
  freq <- stats_malenames[grep(names[i], stats_malenames[,1]), 'freq']
  freq_malenames[i] <- freq*100
}
```

```
#total male and females with certain name
tot_people <- freq_femnames + freq_malenames
#percent of females with name out of all male/female with this name
tot_freq_fem <- freq_femnames/(tot_people)

#bind together percent women with name and associated strength between names and female
#words
fem_name_tot <- as.data.frame(cbind(tot_freq_fem, as.data.frame(assoc_fem_names)))

#plot strength vs. percentage of people with a given name that are women
ggplot(data = fem_name_tot, aes(x = tot_freq_fem*100, y = assoc_fem_names, color=assoc_fem_names)) +
  scale_colour_gradientn(colors = c('blue', 'green','yellow', 'orange','red')) +
  geom_point(size = 3) +
  xlab("Percentage of people with name who are women") +
  ylab("Strength of association of
       name vector with female gender") +
  ggtitle("People with androgynous names") +
  geom_hline(yintercept = 0) +
  theme(legend.position = "none")
```



```
ggsave("./results/fig/androgynous_names.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#Pearson correlation coefficient
p_fem_name <- cor(fem_name_tot[,1], fem_name_tot[,2], use = "pairwise.complete.obs",
                  method = "pearson")
```

```
p_fem_name
```

```
## [1] 0.8117238
```