module_09.R

gregfoster

2021-03-13

```r
#Setting the working directory
setwd("~/Documents/Current_Classes/R_Programming/Open Source R Directory")

#Importing the 2011 Canadian Election Study on Attitudes Towards Abortion
data <- read.csv("can_election.csv")

#Taking a quick look at the data
class(data)
```

```
## [1] "data.frame"
```

```r
summary(data)
```

```
##        X                id            province           population
##  Min.   :   1.0   Min.   :   2.0   Length:2231        Min.   : 105780
##  1st Qu.: 558.5   1st Qu.: 858.5   Class :character   1st Qu.:2515180
##  Median :1116.0   Median :1680.0   Mode  :character   Median :5996930
##  Mean   :1116.0   Mean   :1698.9                      Mean   :5290062
##  3rd Qu.:1673.5   3rd Qu.:2550.0                      3rd Qu.:9439960
##  Max.   :2231.0   Max.   :3455.0                      Max.   :9439960
##      weight          gender            abortion          importance
##  Min.   :  435.3   Length:2231        Length:2231        Length:2231
##  1st Qu.: 4488.8   Class :character   Class :character   Class :character
##  Median : 6153.9   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 7182.2
##  3rd Qu.: 8977.6
##  Max.   :43515.2
##    education           urban
##  Length:2231        Length:2231
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
#The abortion column is in a yes and no format, so this will make a simple grouping c
olumn
unique(data$abortion)
```

```
## [1] "No"  "Yes"
```

```
data$abortion <- factor(data$abortion, ordered=TRUE, levels=c("No", "Yes"))
class(data$abortion) #Now the abortion column can be read as logical values rather th
an characters
```

```
## [1] "ordered" "factor"
```

```
#The Education Column will also be useful as a grouping factor
unique(data$education)
```

```
## [1] "somePS"    "bachelors" "college"    "higher"    "lessHS"     "HS"
```

```
data$education <- factor(data$education, ordered=TRUE, levels=c("lessHS", "HS", "some
PS", "college", "bachelors", "higher"))
class(data$education)
```

```
## [1] "ordered" "factor"
```

```
#Lastly we will want to do the same to the gender column
unique(data$gender)
```

```
## [1] "Female" "Male"
```

```
data$gender <- factor(data$gender, ordered=TRUE, levels=c("Female", "Male"))
class(data$gender)
```

```
## [1] "ordered" "factor"
```

```
#Now most of the important variables are able to be used for analysis
#Let's make some plots!
#The first plot I will make with base R graphics

#I want a good idea of how gender and education impact opinions on abortion
plot(data$gender, data$abortion) #It looks like gender doesn't play a huge role, the
vast majority of people regardless of gender feel that abortion should not be banned.
```
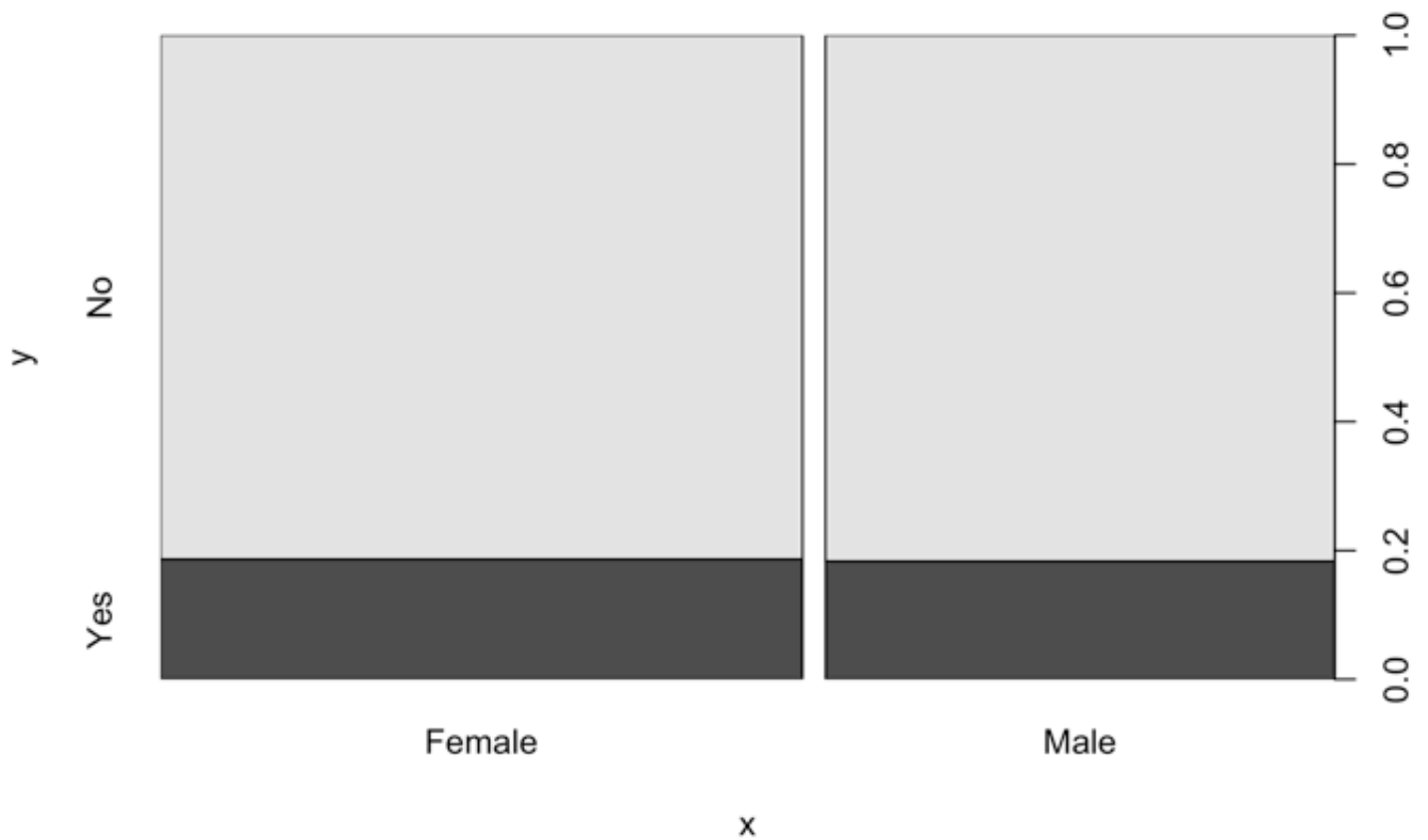
```
plot(data$education, data$abortion)

#The education plot reveals an interesting trend, as the highest educated population,
those with a graduate degree ("higher"), has the lowest percentage of people that thi
nk abortion should be banned.
#Similarly, the least educated population, those that didn't finish high school, has
the highest population of people that think abortions should be banned.

#Now let's see if there is a relationship between Canadian Province and Abortion Sent
iment
unique(data$province)
```
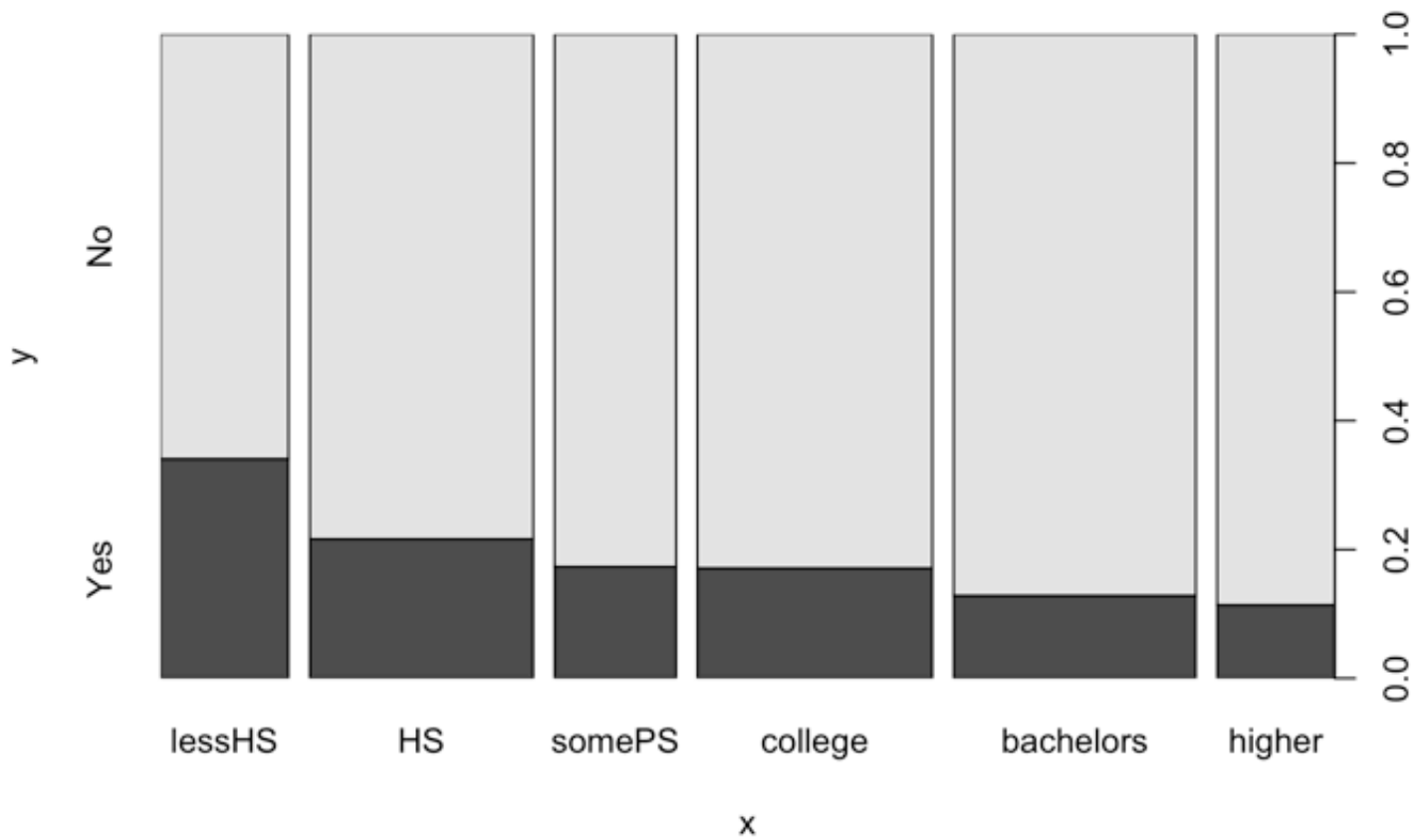
```
##  [1] "BC" "QC" "NL" "ON" "PE" "NS" "NB" "MB" "SK" "AB"
```

```
data$province <- factor(data$province, ordered=TRUE, levels=c("BC", "QC", "NL", "ON",
"PE", "NS", "NB", "MB", "SK", "AB"))
class(data$province)
```

```
## [1] "ordered" "factor"
```
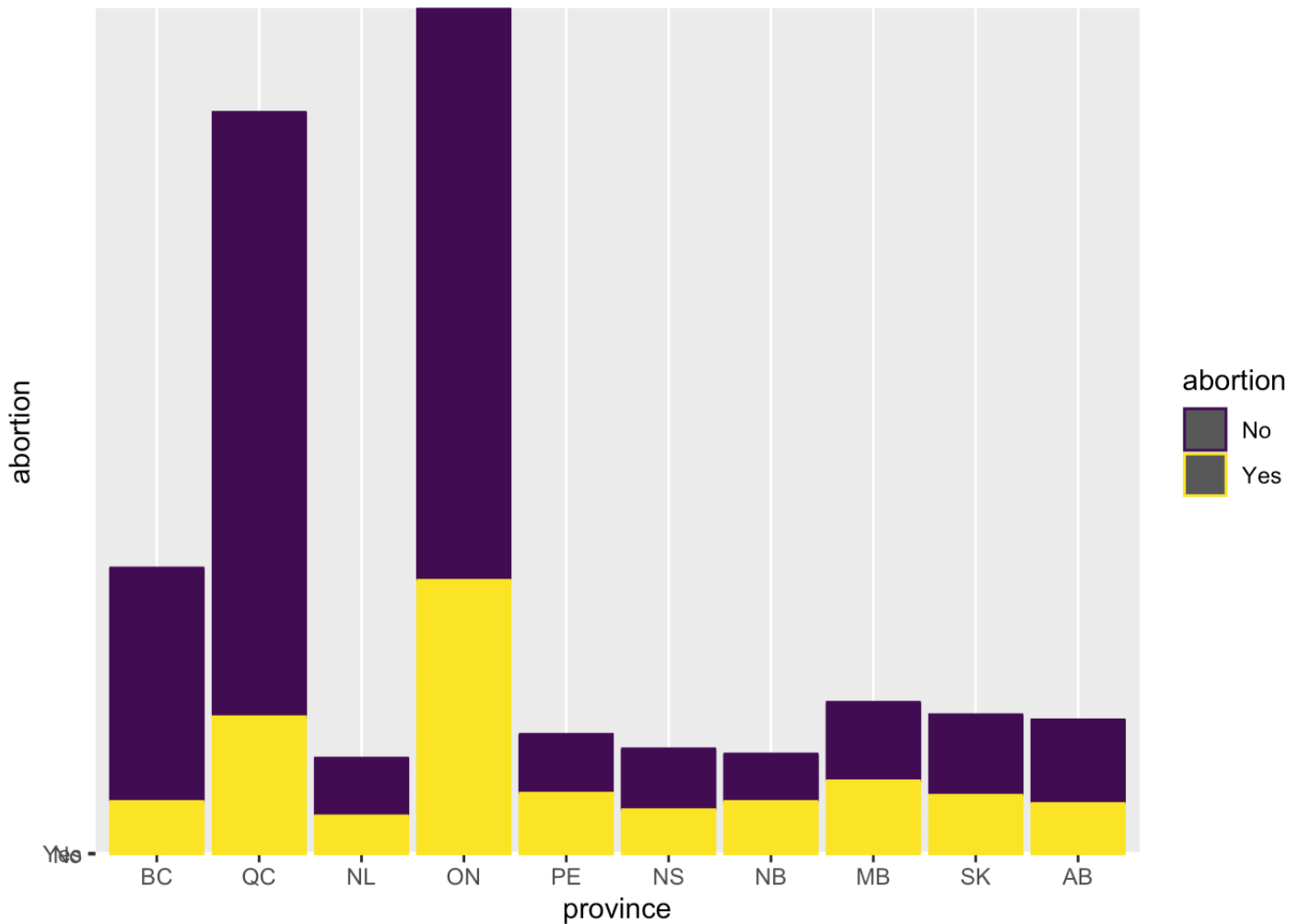
```
library(ggplot2)
```

```r
library(dplyr)
```

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

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

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

```r
ggplot(data, aes(province, abortion, color=abortion)) + geom_col()
```

```
#It looks like Ontario and Quebec have the highest populations in favor of abortion b
eing illegal, highlighted in yellow


#Let's check and see if there is a drastic difference between urban and rural provinc
es
unique(data$urban)
```
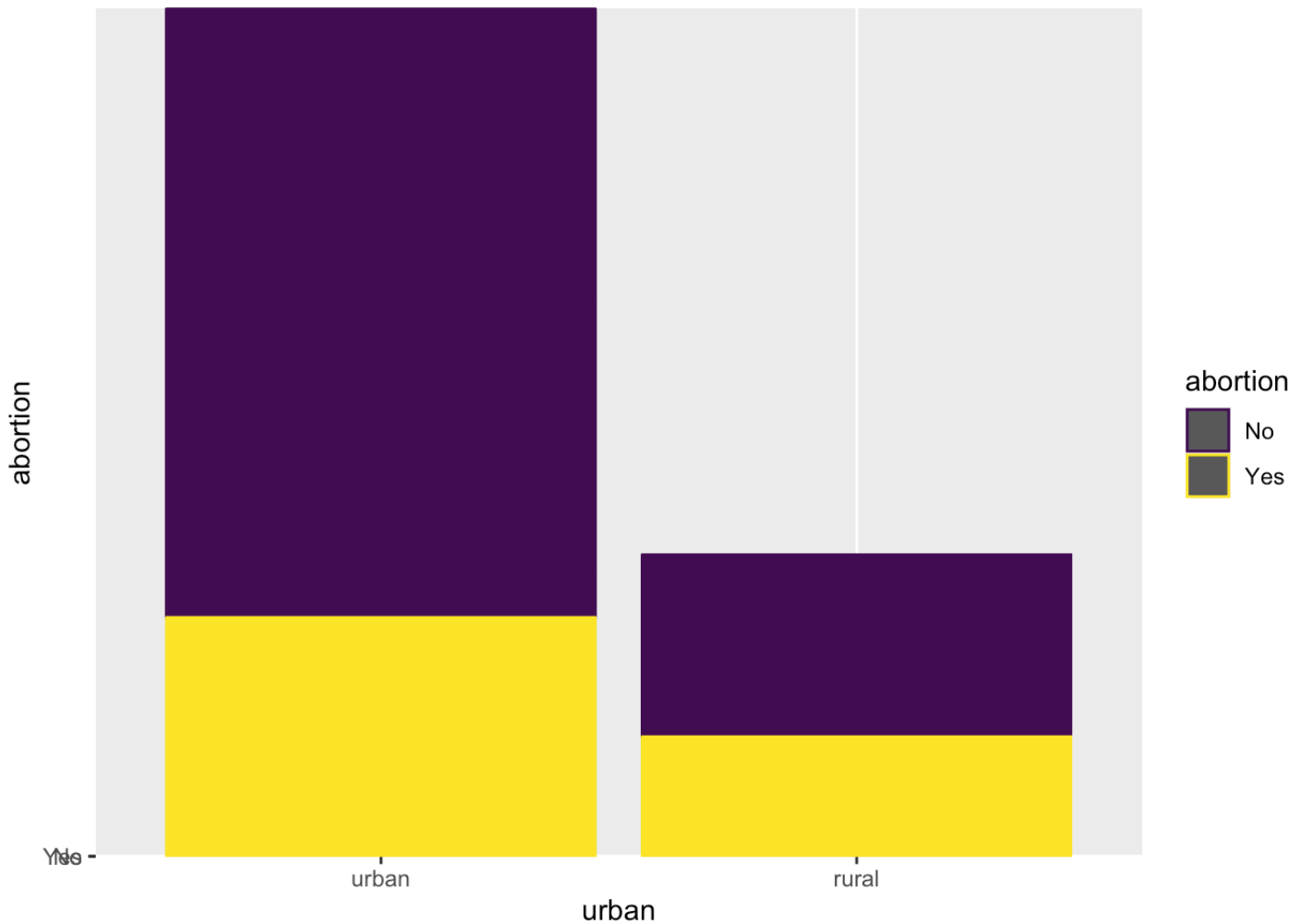
```
## [1] "urban" "rural"
```

```
data$urban <- factor(data$urban, ordered=TRUE, levels=c("urban", "rural"))
class(data$urban)
```

```
## [1] "ordered" "factor"
```

```
ggplot(data, aes(urban, abortion, color=abortion)) + geom_col()
```

```
#It appears that the urban populations have more people overall in favor of getting r
id of abortions, however the rural population has a much higher overall percentage of
people in favor of getting rid of abortions.

#While looking at two variables at a time is well and good, I would like to explore s
everal at a time.
#Let's make a multivariate plot

#First we will have to narrow down our dataset to 3 or 4 variables we want to explore
. I only want province, education, gender, and abortion
colnames(data)
```

```
##  [1] "X"          "id"         "province"   "population" "weight"
##  [6] "gender"     "abortion"   "importance" "education"  "urban"
```

```
data1 <- data[,c(3, 9, 6, 7, 1, 2, 4, 5, 8, 10)]
colnames(data1)
```

```
## [1] "province"   "education"  "gender"     "abortion"   "X"
## [6] "id"         "population" "weight"     "importance" "urban"
```

```
data1 <- data1[,-5:-10]
summary(data1)
```

```
##     province       education      gender      abortion
## ON     :687   lessHS   :267   Female:1244   No :1818
## QC     :652   HS       :467   Male  : 987   Yes: 413
## BC     :252   somePS   :254
## MB     :112   college  :491
## SK     :107   bachelors:506
## AB     :106   higher   :246
## (Other):315
```

```
library(cdparcoord)
```

```
## Loading required package: data.table
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## Loading required package: plotly
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
## Loading required package: freqparcoord
```

```
## Loading required package: parallel
```

```
## Loading required package: GGally
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
## Loading required package: FNN
```

```
## Loading required package: mvtnorm
```

```
##
##
##
##
##
##     For a quick introduction, type ?freqparcoord, and
##     run the examples, making sure to read the comments.
##
##
##
```

```
## Loading required package: partools
```

```
## Loading required package: regtools
```

```
## Loading required package: dummies
```

```
## dummies-1.5.6 provided by Decision Patterns
```

```
## Loading required package: sandwich
```

```
##
##
##
##
##
## ********************
##
##
##
## Latest version of regtools at GitHub.com/matloff
##
##
## Type "?regtools" for function list.
```

```
## Loading required package: pdist
```

```
## Latest version of partools at GitHub.com/matloff
```

```
##
##
##
##
##
## ********************
##
##
##
##
##
##
## Type ?quickstart for cdparcoord quick start
##
##
##
##
## `
```

```
mm <- discretize(data1,nlevels=100)
discparcoord(mm,k=5000,saveCounts=FALSE,name="2011 Canadian Sentiment on Abortions")
```

2011 Canadian Sentiment on Abortions