Q2

Trying to load all 28 CSV documents kept crashing my (rather frail) Lenovo computer - I apologize for not getting this to work, my Macbook Pro is on the way!

Q3

```
## 'data.frame':    43400 obs. of  12 variables:
##  $ id               : int  16523 56543 32257 52800 41413 15266 28674 10460 63884 37893 ...
##  $ gender           : Factor w/ 5 levels "female","Female",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ age              : num  8 70 47 52 75 32 74 79 37 37 ...
##  $ hypertension     : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ heart_disease    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ ever_married     : Factor w/ 2 levels "No","Yes": 1 2 2 2 2 2 2 2 2 2 ...
##  $ work_type        : Factor w/ 5 levels "children","Govt_job",..: 4 4 4 4 5 4 5 2 4 4 ...
##  $ Residence_type   : Factor w/ 2 levels "Rural","Urban": 2 1 2 2 1 1 2 2 1 1 ...
##  $ avg_glucose_level: num  110.9 69 210.9 77.6 243.5 ...
##  $ bmi              : num  17.6 35.9 50.1 17.7 27 32.3 54.6 35 39.4 26.1 ...
##  $ smoking_status   : Factor w/ 4 levels "","formerly smoked",..: 1 2 1 2 3 4 3 1 3 2 ...
##  $ stroke           : int  0 0 0 0 0 0 0 0 0 0 ...

## 'data.frame':    43400 obs. of  12 variables:
##  $ id               : int  16523 56543 32257 52800 41413 15266 28674 10460 63884 37893 ...
##  $ gender           : chr  "female" "female" "female" "female" ...
##  $ age              : num  8 70 47 52 75 32 74 79 37 37 ...
##  $ hypertension     : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ heart_disease    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ ever_married     : chr  "No" "Yes" "Yes" "Yes" ...
##  $ work_type        : chr  "Private" "Private" "Private" "Private" ...
##  $ Residence_type   : chr  "Urban" "Rural" "Urban" "Urban" ...
##  $ avg_glucose_level: num  110.9 69 210.9 77.6 243.5 ...
##  $ bmi              : num  17.6 35.9 50.1 17.7 27 32.3 54.6 35 39.4 26.1 ...
##  $ smoking_status   : chr  "" "formerly smoked" "" "formerly smoked" ...
##  $ stroke           : int  0 0 0 0 0 0 0 0 0 0 ...

## 'data.frame':    43400 obs. of  12 variables:
##  $ id               : int  16523 56543 32257 52800 41413 15266 28674 10460 63884 37893 ...
##  $ gender           : Factor w/ 5 levels "female","Female",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ age              : num  8 70 47 52 75 32 74 79 37 37 ...
##  $ hypertension     : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ heart_disease    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ ever_married     : Factor w/ 2 levels "No","Yes": 1 2 2 2 2 2 2 2 2 2 ...
##  $ work_type        : Factor w/ 5 levels "children","Govt_job",..: 4 4 4 4 5 4 5 2 4 4 ...
##  $ Residence_type   : Factor w/ 2 levels "Rural","Urban": 2 1 2 2 1 1 2 2 1 1 ...
##  $ avg_glucose_level: num  110.9 69 210.9 77.6 243.5 ...
##  $ bmi              : num  17.6 35.9 50.1 17.7 27 32.3 54.6 35 39.4 26.1 ...
##  $ smoking_status   : Factor w/ 4 levels "","formerly smoked",..: 1 2 1 2 3 4 3 1 3 2 ...
##  $ stroke           : int  0 0 0 0 0 0 0 0 0 0 ...
```
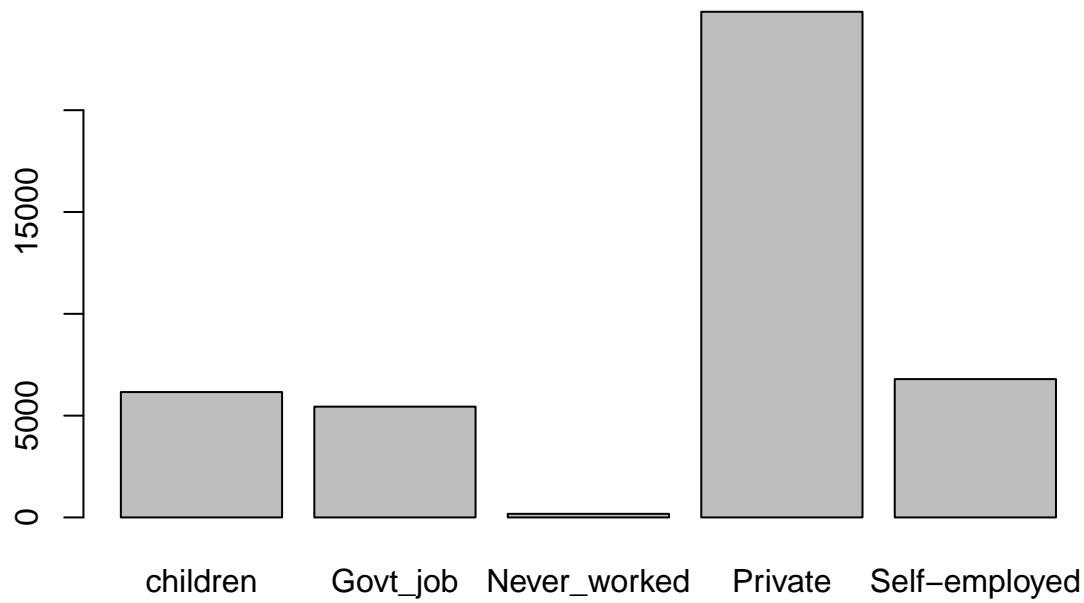
```r
levels(heartAttack_1$work_type)
```

```
## [1] "children"      "Govt_job"      "Never_worked"  "Private"
## [5] "Self-employed"
```

```r
# min(heartAttack_1$work_type) # not ordered

barplot(table(heartAttack_1$work_type))

heartAttack_1$work_type <- ordered(heartAttack_1$work_type)

barplot(table(heartAttack_1$work_type))
```
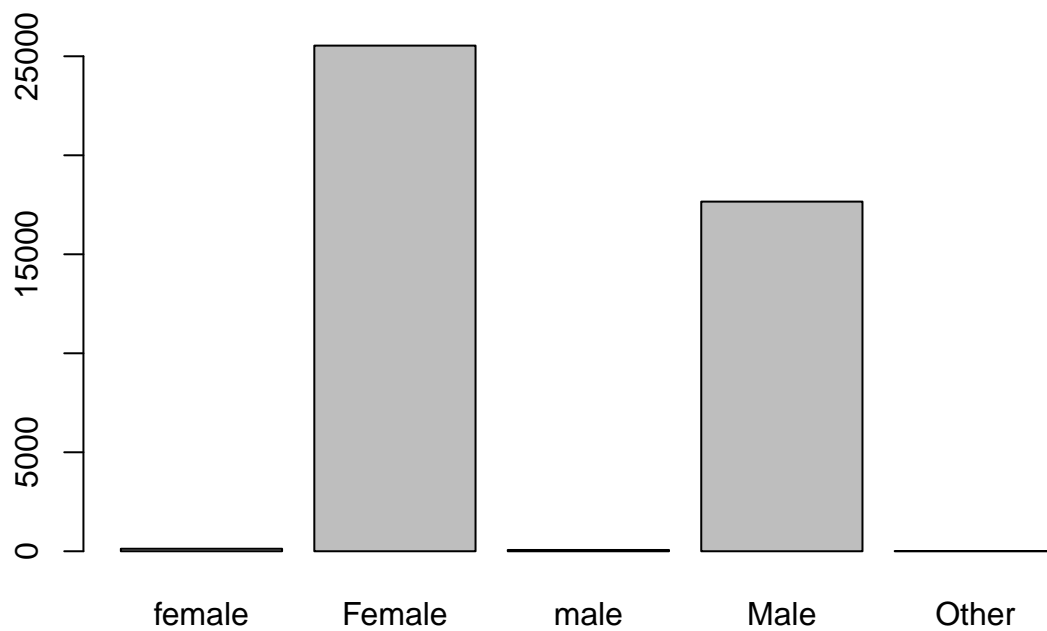
```
barplot(table(heartAttack_1$gender)) # there are spelling differences that create 5 factors instead of
```
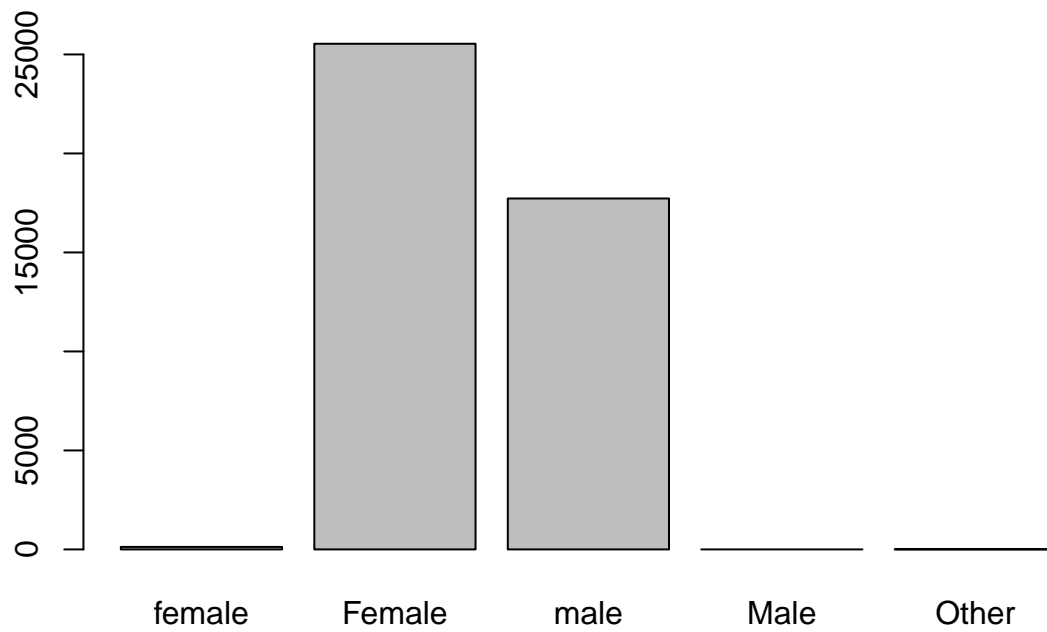


```
heartAttack_1$gender[heartAttack_1$gender == "Male"] <- "male"

barplot(table(heartAttack_1$gender)) # the factor still exists, even though there are 0 observations in
```

```
heartAttack_1$gender[heartAttack_1$gender == "Female"] <- "female"

table(droplevels(heartAttack_1)$gender)
```

```
##
## female   male  Other
##  25665  17724     11
```

Q4

Create a vector of 10,000 random variates $\sim \mathcal{U}\{10^{-15}, 10^5\}$. These numbers represent wavelengths (in meters) of photons hitting the Hubble telescope.

```
wavelengthsNums <- exp(runif(10000,log(1e-15),log(1e5)))
wavelengths <- exp(runif(10000,log(1e-15),log(1e5)))
```

- Convert this vector into an ordinal factor with levels as the type of wave. Levels should be ordered in terms of increasing wavelength.

```
library("ggplot2")
gamma <- 1e-11
xray <- 10e-8
uv <- 4 * 10e-7
visible <- 7.5 * 10e-7
infrared <- 10e-6
microwave <- 10e-3
radio <- 10e3

frequencies <- c(gamma, xray, uv, visible, infrared, microwave, radio)

wavelengths <- cut(wavelengths,
                   breaks=c(-Inf, gamma, xray, uv, visible, infrared, microwave, radio, Inf),
                   ordered_result = TRUE,
                   right=TRUE,
                   labels=c("gamma", "xray", "uv", "visible", "infrared", "microwave", "radio", "beyond
                   )
```

- Create a boxplot for each factor level.

```
ggplot(data.frame(wavelengths, wavelengthsNums), aes(x=wavelengths, y=log(wavelengthsNums), fill=wavelen
  geom_boxplot() +
  labs(title="Plot of Wavelengths and Log Wavelength ")
```

Plot of Wavelengths and Log Wavelength
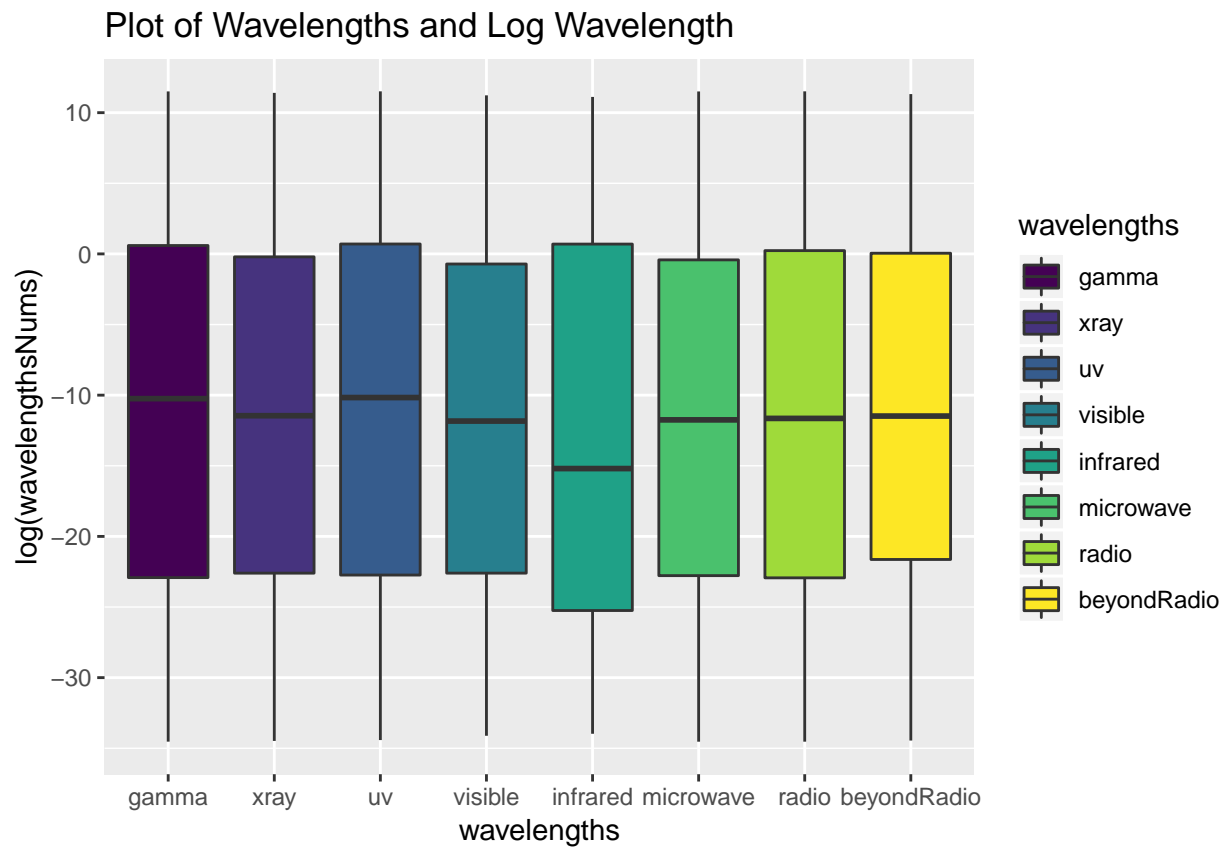
- How many photons can you see with the the naked eye?

```
10000 * ((visible-uv)/(10e5-10e-15))
```

```
## [1] 3.5e-08
```