R coding for public policy

Final Part I

[Name removed]

Instruction: follow the steps to create a function that takes a name of the county and a year and outputs the most popular baby boy and baby girl name.

Dataset: The Baby names dataset from lecture 3 link: https://health.data.ny.gov/Health/Baby-Names-Beginning-2007/jxy9-yhdk *make sure you find the correct url for the csv file

#import and clean baby name dataset, correct the column classes,remove the missing rows and rename the columns for your convenience.

```
baby<-read.csv("https://health.data.ny.gov/api/views/jxy9-yhdk/rows.csv?accessType=DOWNLOAD")


a<-sapply(baby, is.numeric)
a

##       Year First.Name     County        Sex      Count
##       TRUE      FALSE      FALSE      FALSE       TRUE

baby[, !a]<-lapply(baby[, !a], toupper)
baby<-na.omit(baby)
```

#refer to following code for assignment 3 Problem 3 [Baby] use max() to find the most popular baby boy name and the most popular baby girl name for “KINGs” county in 2016 attach(baby) baby3<-baby[County==“KINGS” & Year==2016,] b<-max(baby3[baby3$Sex==“M”,]$Count)baby3[baby3Count==b & baby3$Sex==“M”,]

```
    g<-max(baby3[baby3$Sex=="F",]$Count)
    baby3[baby3$Count==g & baby3$Sex="F",]
    detach(baby)
```

#1. which part of the code attaches the dataset to the environment, so that we can use column names directly

```
attach(baby)
```

#2. which part of the code produces a subset of data with a specific county and year?

```
baby3<-baby[County=="KINGS" & Year==2016, ]
```

#3. which part of the code FINDS the position for the most popular baby boy name?

```
b<-max(baby3[baby3$Sex=="M",]$Count)
b

## [1] 231
```

#4. which part of the code PRINTS the most popular baby boy name?

```
baby3[baby3$Count==b & baby3$Sex=="M",]

##   Year First.Name County Sex Count
## 1 2016      DAVID  KINGS   M   231
```

#5. which part of the code FINDS the position for the most popular baby girl name?

```
g<-max(baby3[baby3$Sex=="F",]$Count)
```

#6. which part of the code PRINTS the most popular baby girl name?

```
baby3[baby3$Count==g & baby3$Sex=="F",]

##   Year First.Name County Sex Count
## 5 2016     OLIVIA  KINGS   F   210
```

#7. which part of the code detaches the dataset?

```
detach(baby)
```

#write the function script below step1. name the function most_popular_names step2. call function(){} step3. start programming within {} step3.1 attach the dataset baby step3.2 subset the dataset baby with "KINGS" and 2016 step3.3 find the position for the most popular baby boy name step3.4 find the position for the most popular baby girl name step3.5 create a list called "output" with the two pieces of codes that print the most popular names step3.6 return print(output) you can use return(print(output)) or print(output) step3.7 detach the data baby step3.8 replace the word baby with data, "KINGS" with c, 2016 with y Step 4. fill out the formals with data, c, y Step 5. Run the function script

```
most_popular_names<- function(data,c,y){
        attach(data)
        data3<-data[County== c & Year== y, ]
        b<-max(data3[data3$Sex=="M",]$Count)
        g<-max(data3[data3$Sex=="F",]$Count)
        boy<-data3[data3$Count==b & data3$Sex=="M",]
        girl<-data3[data3$Count==g & data3$Sex =="F",]
        output<-list(boy,girl)
        return(print(output))
        detach(data)
      }
```

Step 6. test the function: what are the most popular names in BRONX in 2015

```
most_popular_names(baby,"BRONX", 2015)
```

```
## [[1]]
##       Year First.Name County Sex Count
## 89952 2015       LIAM  BRONX   M   184
##
## [[2]]
##       Year First.Name County Sex Count
## 89994 2015     SOPHIA  BRONX   F   121
```