

MTH-245 Final project Part 2
Fall 2022

Name:

```
library("tidyverse")
library("xtable")
library("patchwork")
dat.births <- read_csv("/Users/marissapatel/Downloads/NCbirths.csv") %>% dplyr::select(BirthWeightGm, Race, Sex,
  mutate( SexCat = case_when( Sex == 1 ~ "male",
                                Sex == 2 ~ "female")) %>%
  mutate( SmokeCat = case_when( Smoke == 1 ~ "yes",
                                Smoke == 0 ~ "no"))

## Error:  '/Users/marissapatel/Downloads/NCbirths.csv' does not exist.
```

1 Part 1: Abstract

2 Part 2: Introduction

3 Part 3: Exploratory Data Analysis

a. Graphically summarize the variables in your dataset.

```
violin.BirthWeight <- ggplot(dat.births, aes(x=BirthWeightGm, y=""))+
  geom_violin(fill = "lightblue",
              trim = FALSE)+
  geom_boxplot(width = .3,
               fill = "white") +
  theme_bw()+
  xlab("Birth Weights")+
  ylab(" ")+
  ggtitle("Distribution of Birth Weights",
          subtitle = "NCBirths Data")

violin.GestationPeriod <- ggplot(dat.births, aes(x=Weeks, y=""))+
  geom_violin(fill = "lightblue",
              trim = FALSE)+
  geom_boxplot(width = .3,
               fill = "white") +
  theme_bw()+
  xlab("Weeks")+
  ylab(" ")+
  ggtitle("Distribution of Gestation Period",
          subtitle = "NCBirths Data")

violin.MomAge <- ggplot(dat.births, aes(x=MomAge, y=""))+
  geom_violin(fill = "lightblue",
              trim = FALSE)+
  geom_boxplot(width = .3,
               fill = "white") +
  theme_bw()+
  xlab("Age (years)")+
```

```

ylab(" ") +
ggtitle("Distribution of Mothers' Ages",
        subtitle = "NCBirths Data")

violin.Race <- ggplot(dat.births, aes(x=BirthWeightGm, y=MomRace)) +
  geom_violin(fill = "lightblue",
             trim = FALSE) +
  geom_boxplot(width = .3,
              fill = "white") +
  theme_bw() +
  xlab("Birth Weight") +
  ylab("Races") +
  ggtitle("Distribution of Birth Weights by Mothers' Race",
          subtitle = "NCBirths Data")

violin.Sex <- ggplot(dat.births, aes(x=BirthWeightGm, y=SexCat)) +
  geom_violin(fill = "lightblue",
             trim = FALSE) +
  geom_boxplot(width = .3,
              fill = "white") +
  theme_bw() +
  xlab("Birth Weight") +
  ylab("Sex") +
  ggtitle("Distribution of Birth Weights by Childrens' Sex",
          subtitle = "NCBirths Data")

violin.Smoke <- ggplot(dat.births, aes(x=BirthWeightGm, y=SmokeCat)) +
  geom_violin(fill = "lightblue",
             trim = FALSE) +
  geom_boxplot(width = .3,
              fill = "white") +
  theme_bw() +
  xlab("Birth Weights") +
  ylab("Smokes (yes or no)") +
  ggtitle("Distribution of Birth Weights by Mothers Who Smoke",
          subtitle = "NCBirths Data")

violin.BirthWeight + violin.GestationPeriod + violin.MomAge /
  violin.Race + violin.Sex + violin.Smoke

```

```

## Error in ggplot(dat.births, aes(x = BirthWeightGm, y = "")): object 'dat.births' not found
## Error in ggplot(dat.births, aes(x = Weeks, y = "")): object 'dat.births' not found
## Error in ggplot(dat.births, aes(x = MomAge, y = "")): object 'dat.births' not found
## Error in ggplot(dat.births, aes(x = BirthWeightGm, y = MomRace)): object 'dat.births' not
found
## Error in ggplot(dat.births, aes(x = BirthWeightGm, y = SexCat)): object 'dat.births' not
found
## Error in ggplot(dat.births, aes(x = BirthWeightGm, y = SmokeCat)): object 'dat.births' not
found
## Error in eval(expr, envir, enclos): object 'violin.BirthWeight' not found

```

Figure 1: Violin plots of each variable.

the shape of the distribution of each variable any unusual looking observations

From 1 we can see that there is variability in almost all of the variables have and many odd observations.

```

histogram.BirthWeight<- ggplot(dat.births, aes(x=BirthWeightGm))+
  geom_histogram(fill = "lightblue",
                 color = "black",
                 bins = 5) +
  theme_bw() +
  xlab("Birth Weights")+
  ylab("Count of Weight(gm)")+
  ggtitle("Frequencies of Birth Weights")

histogram.Gestation<- ggplot(dat.births, aes(x=Weeks))+
  geom_histogram(fill = "lightblue",
                 color = "black",
                 bins = 5) +
  theme_bw() +
  xlab("Gestation Period")+
  ylab("Count of Weeks")+
  ggtitle("Frequencies of Gestation Periods")

histogram.MomAge <- ggplot(dat.births, aes(x=MomAge))+
  geom_histogram(fill = "lightblue",
                 color = "black",
                 bins = 5) +
  theme_bw() +
  xlab("Ages of Mothers(years)")+
  ylab("Count of Ages")+
  ggtitle("Frequencies of Ages")

histogram.BirthWeight + histogram.Gestation + histogram.MomAge

```

```

## Error in ggplot(dat.births, aes(x = BirthWeightGm)): object 'dat.births' not found
## Error in ggplot(dat.births, aes(x = Weeks)): object 'dat.births' not found
## Error in ggplot(dat.births, aes(x = MomAge)): object 'dat.births' not found
## Error in eval(expr, envir, enclos): object 'histogram.BirthWeight' not found

```

Figure 2: Grid of histograms for the quantitative variables.

2 shows that the quantitative variables do not follow normal distributions and are all skewed.

b. Numerically summarize the variables in your dataset.

```

sumstats <- dat.births %>% summarize(mean=mean(diff.happ),
                                     variance=var(diff.happ),
                                     sample_size = n())

## Error in summarize(., mean = mean(diff.happ), variance = var(diff.happ), : object 'dat.births'
not found

```

c. Create a scatterplot matrix and table of correlations.

```

library(GGally)
correlationsmatrix <- ggpairs(dat.births, columns = c(1,2, 4:7))

correlationsmatrix

```

```
## Error in library(GGally): there is no package called 'GGally'  
## Error in ggpairs(dat.births, columns = c(1, 2, 4:7)): could not find function "ggpairs"  
## Error in eval(expr, envir, enclos): object 'correlationsmatrix' not found
```

Figure 3: Matrix of ScatterPlots and Correlations for the variables.

d. **Other interesting plots.**

Plot significant correlations - weeks and birth weight

e. **Comment on...**

the shape of the distribution of each variable

the relationship between the response and the quantitative predictors

any unusual looking observations

any other interesting takeaways