

```

---
title: "chapter 5 Summarization Lab B"
author: "Your Name Here"
date: ""
output: html_document
---

```

##### Remember to change the `author:` field on this Rmd file to your own name.

We'll begin by doing all the same data processing as in lecture.

```

```{r}
library(MASS)

# Assign more descriptive variable names
colnames(birthwt) <- c("birthwt.below.2500", "mother.age", "mother.weight",
  "race", "mother.smokes", "previous.prem.labor", "hypertension", "uterine.irr",
  "physician.visits", "birthwt.grams")

# Assign more descriptive factor levels and convert variables to factors as needed
library(plyr)
birthwt <- transform(birthwt,
  race = as.factor(mapvalues(race, c(1, 2, 3),
    c("white", "black", "other"))),
  mother.smokes = as.factor(mapvalues(mother.smokes,
    c(0,1), c("no", "yes"))),
  hypertension = as.factor(mapvalues(hypertension,
    c(0,1), c("no", "yes"))),
  uterine.irr = as.factor(mapvalues(uterine.irr,
    c(0,1), c("no", "yes"))),
  birthwt.below.2500 = as.factor(mapvalues(birthwt.below.2500,
    c(0,1), c("no", "yes")))
)
```

```

#### 1. aggregate() vs tapply()

One of the advantages of aggregate() is that it makes it easier to view summary tables when grouping on more than two factors.

**\*\* (a) \*\*** Use the `tapply()` function to calculate mean `birthwt.grams` grouped by race, mother's smoking status, and hypertension.

```

```{r}
# Edit me
```

```

**\*\* (b) \*\*** Repeat part (a), this time using the `aggregate()` function.

```

```{r}
# Edit me
```

```

#### 2. by() practice

In today's lecture we used the `by()` function to calculate the correlation between mother's age and birth weight broken down by mother's smoking status. Here's the command that was used:

```

```{r}
by(data = birthwt[c("birthwt.grams", "mother.age")],
  INDICES = birthwt["mother.smokes"],
  FUN = function(x) {cor(x[,1], x[,2])})
```

```

**\*\* (a) \*\*** Use the `by()` function to calculate the correlation between mother's age and birth weight broken down by race.

```

```{r}
# Edit me
```

```

**\*\* (b) \*\*** Does the correlation appear to vary by race?

### ### 3. Scatterplot practice

In class we made a pretty nice scatterplot using the following command.

```
```{r, fig.align='center', eval=FALSE}
# Make sure that the color-blind palette is defined
cbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00",
"#CC79A7")

points.base <- c(16, 17) # 16: circle, 17: triangle
cols.base <- cbPalette[2:3] # blue and orange points
point.symbols <- points.base[birthwt$mother.smokes]
point.cols <- cols.base[birthwt$mother.smokes]
with(birthwt, plot(mother.age, birthwt.grams,
                  main="Birth Weight by Mother's Age",
                  xlab="Mother's Age (Years)",
                  ylab="Birth Weight (Grams)",
                  col=point.cols,
                  pch=point.symbols))
legend("bottomright", c("Nonsmoker", "Smoker"), col=cols.base, pch=points.base)
```
```

**(a)** Try to create a similar plot, but instead of using different colors and symbols based on the value of `mother.smokes`, do so based on the value of `race`.

```
```{r, fig.height=6, fig.width=6, fig.align='center'}
# Edit me
```
```

**(b)** Use the `abline` command to add a horizontal line to the plot at `birthwt.grams = 2500`.

```
```{r, fig.height=6, fig.width=6, fig.align='center'}
# Edit me
```
```