Real World Example 2

Lisa Gandolfi

R Problems

Use the `states20` data set to identify potential explanations for state-to-state differences in infant mortality (`states20$infant_mort`).

1. Generate a scatterplot showing the relationship between per capita income (`states20$PCincome2020`) and infant mortality in the states.

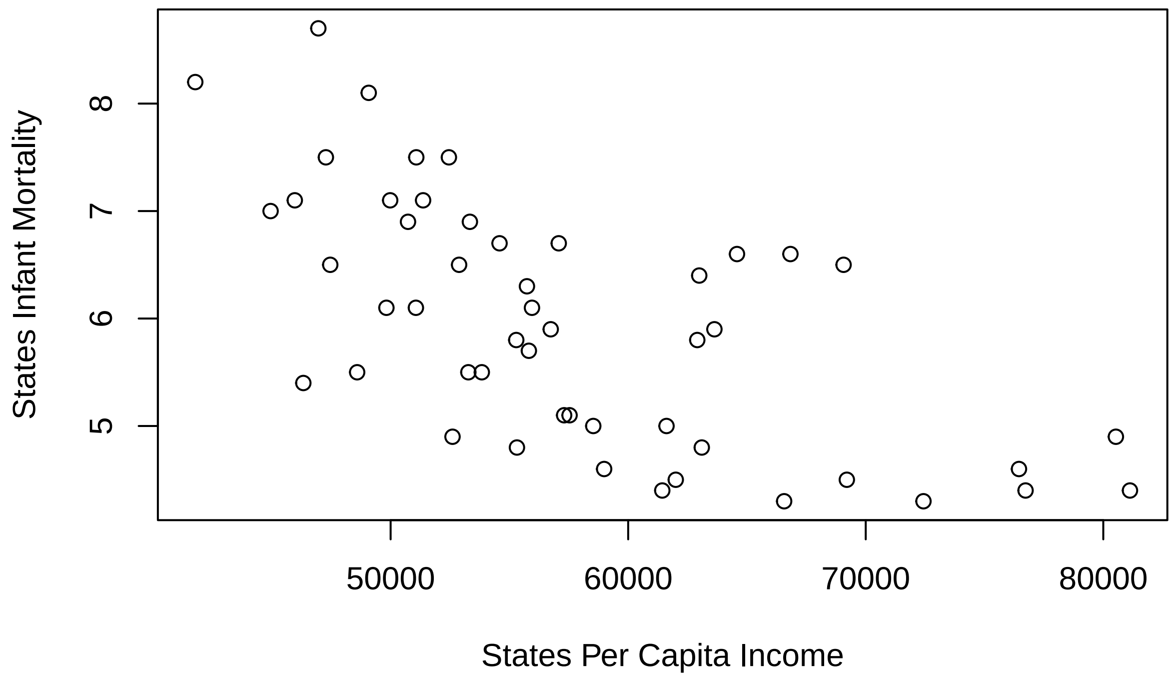
```
library(descr)
library(DescTools)
library(Hmisc)
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:DescTools':
##
##     %nin%, Label, Mean, Quantile

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(ppcor)
```

```
## Loading required package: MASS
```

```
library(readxl)
load("data/states20.rda")

#scatterplot between states20$PCincome2020 and states20$infant_mort, indep, dep
plot(states20$PCincome2020,states20$infant_mort,
xlab="States Per Capita Income",
ylab="States Infant Mortality")
```



The scatterplot shows that there seems to be a fairly strong relationship between states' infant mortality and per-capita income, since the upper right and lower left corners result to be empty. As a result, we can say that there seems to be a pretty strong, negative relationship between infant mortality and per-capita income. However, the observations are not very clustered, especially as per-capita income rises, which suggests that in order to finalize the directional strength of the relationship, we should look into Pearson's R. In addition, we might be overestimating the relationship between these two variables since we are not taking into consideration other important variables.

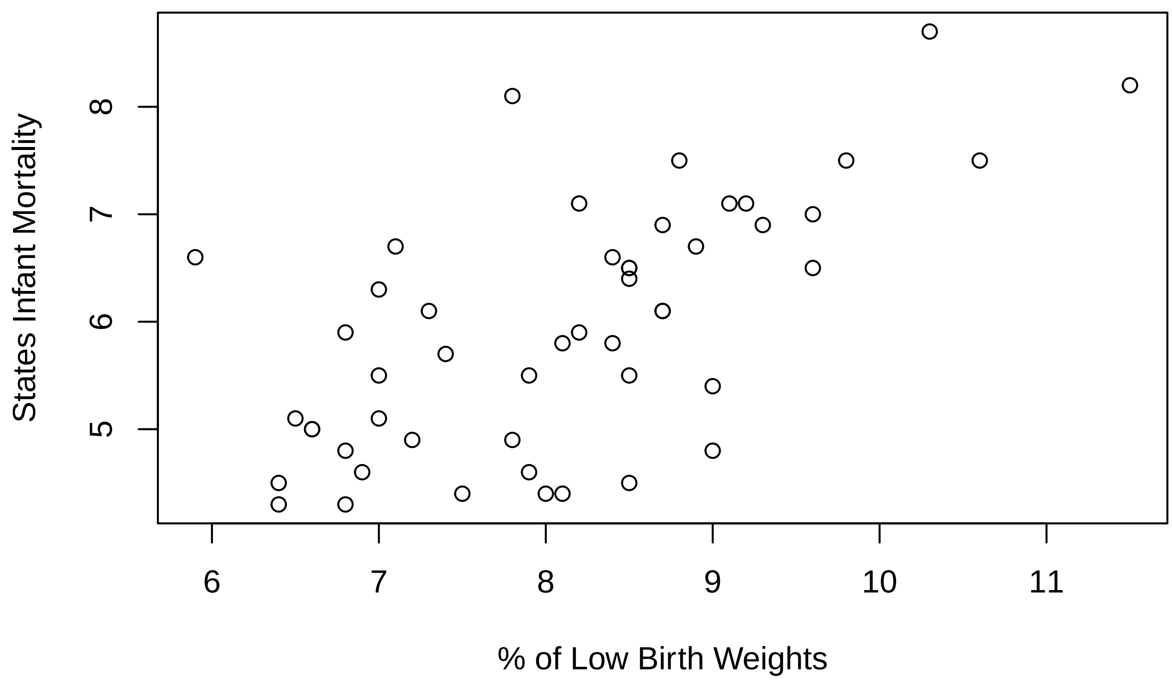
```r
#correlation test
cor.test(states20$infant_mort,states20$PCincome2020)
```

```
##
##  Pearson's product-moment correlation
##
## data:  states20$infant_mort and states20$PCincome2020
## t = -5.9177, df = 48, p-value = 3.349e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.7857705 -0.4530238
## sample estimates:
##        cor
## -0.6494741
```

The results from Pearson's R show a strong, negative relationship (-0.64) between states' infant mortality and per-capita income. This confirms our initial findings from the scatterplot interpretation, and since Pearson's R results can range from 0 (no relationship to +/-1 (strong positive or negative relationship), we can infer that this relationship is quite strong. In addition, the p-value for this relationship is 3.349e-07, which is very close to zero. Therefore, we can reject the null hypothesis and conclude that the true correlation is not equal to zero.

2. Now use `lowbirthwt` (% of births that are technically low birth weight) as the independent variable.

```r
#scatterplot between states20$lowbirthwt and states20$infant_mort, indep, dep
plot(states20$lowbirthwt,states20$infant_mort,
xlab="% of Low Birth Weights",
ylab="States Infant Mortality")
```



The scatterplot shows that there seems to be a fairly strong positive relationship between states' % of low birth weights and infant mortality, since the lower right and upper left corners result to be empty. As a result, we can say that there seems to be a pretty strong, positive relationship between % of low birth weights and infant mortality. However, the observations are not tightly clustered although they look more clustered than the previous graph, especially percentage of low birth weights rises, which suggests that in order to finalize the directional strength of the relationship, we should look into Pearson's R. Nevertheless, we might be mistakingly estimating the relationship between these two variables since we are not taking into consideration other crucial variables.

```r
#correlation test
cor.test(states20$infant_mort,states20$lowbirthwt)
```

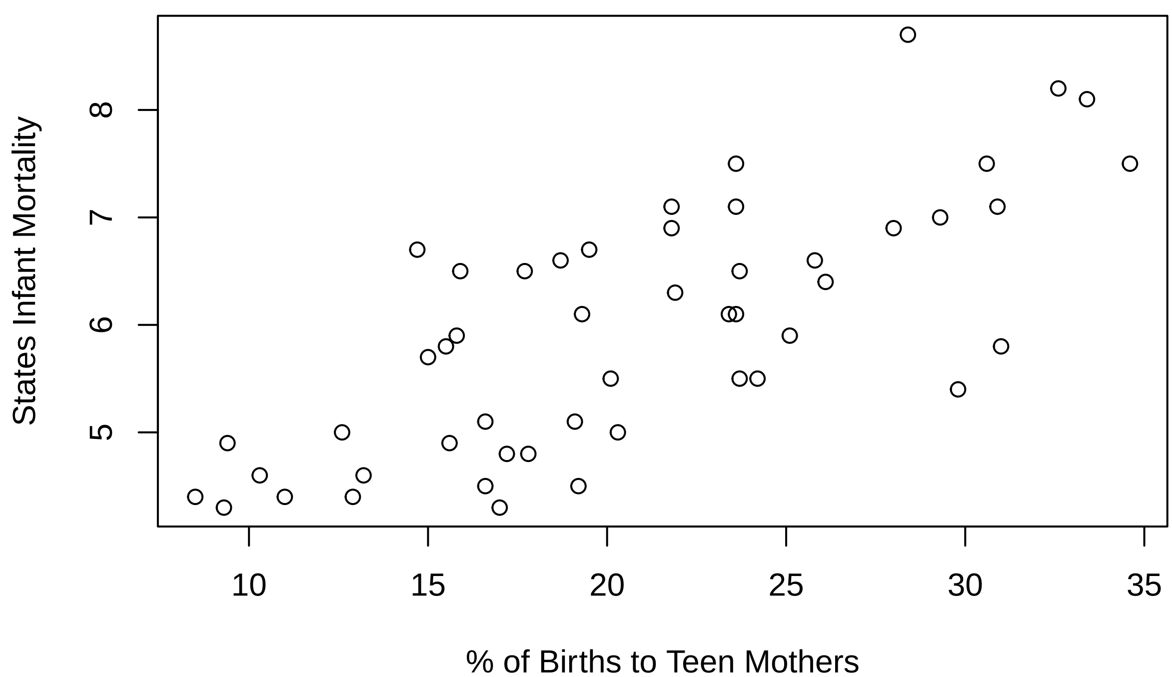
```
##
##  Pearson's product-moment correlation
##
```

```
## data:  states20$infant_mort and states20$lowbirthwt
## t = 5.7918, df = 48, p-value = 5.206e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4419266 0.7804039
## sample estimates:
##       cor
## 0.6413788
```

The results from Pearson's R show a strong, positive relationship (0.64) between states' infant mortality and percentage of low birth weights. This confirms our initial findings from the scatterplot interpretation, and since Pearson's R results can range from 0 (no relationship to +/-1 (strong positive or negative relationship), we can infer that this relationship is quite strong. In addition, the p-value for this relationship is 5.206e-07, which is very close to zero. Therefore, we can reject the null hypothesis and conclude that the true correlation is not equal to zero.

3. Now use `teenbirth` (the % of births to teen mothers) as the independent variable.

```
#scatterplot between states20$teenbirth and states20$infant_mort, indep, dep
plot(states20$teenbirth,states20$infant_mort,
xlab="% of Births to Teen Mothers",
ylab="States Infant Mortality")
```



The scatterplot shows that there seems to be a strong positive relationship between states' % of births to teen mothers and infant mortality, since the lower right and upper left corners result to be empty, as in the previous graph. As a result, we can say that there seems to be a pretty strong, positive relationship between

% of births to teen mothers and infant mortality. However, the observations are not tightly clustered, and look fairly less clustered than the previous graphs. However, the relationship between the two variables seems to be far more linear than previous ones, since all the observations seem to be following a linear relationship and do not present any extreme values. As percentage of births to teen mothers rises, the values seem more scattered, which suggests that in order to finalize the directional strength of the relationship, we should look into Pearson's R. Nevertheless, we might be wrongly estimating the relationship between these two variables since we are not taking into consideration other crucial variables.

```
#correlation test
cor.test(states20$infant_mort,states20$teenbirth)
```

```
##
##  Pearson's product-moment correlation
##
## data:  states20$infant_mort and states20$teenbirth
## t = 7.5626, df = 48, p-value = 1.015e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5775469 0.8427400
## sample estimates:
##       cor
## 0.7373578
```

The results from Pearson's R show a very strong, positive relationship (0.73) between states' infant mortality and percentage of births to teen mothers. This confirms our initial findings from the scatterplot interpretation, and since Pearson's R results can range from 0 (no relationship to +/-1 (strong positive or negative relationship), we can infer that this relationship is very strong, far more than the previous ones. In addition, the p-value for this relationship is 1.015e-09, which is very close to zero. Therefore, we can reject the null hypothesis and conclude that the true correlation is not equal to zero.

4. Now use an independent variable of your choosing from the `states20` data set that you expect to be related to infant mortality in the states.

```
#scatterplot between states20$abortion_laws and states20$infant_mort, indep, dep
plot(states20$abortion_laws,states20$infant_mort,
xlab="# of Abortion Restriction in State Laws",
ylab="States Infant Mortality")
```

For this exercise, I decided to use the variable "abortion_laws", which is the number of abortion laws within the state's legislature. I thought that it would be interesting to investigate the relationship between this variable and infant mortality, due to the prominence that abortion laws have had recently in the news. We might imagine that a state with a high number of abortion laws might either induce to a higher or lower state infant mortality, especially if that means that young mothers cannot decide to terminate pregnancies. The scatterplot shows that there does not seem to be a strong relationship between the Number of Abortion Laws with a State's Legislature and infant mortality. However, we can observe that the upper left and lower right corners still seem to be mostly empty, thus suggesting a somewhat positive relationship. The observations are also not tightly clustered as previous graphs. Differently from previous graphs, many results seem to be stacked and similar numbers of abortion laws yield diverse infant mortality rates. In order to finalize the directional strength of the relationship, we should look into Pearson's R. Nevertheless, we might be wrongly estimating the relationship between these two variables since we are not taking into consideration other crucial variables.

```
#correlation test
cor.test(states20$infant_mort,states20$abortion_laws)
```

```
##
##  Pearson's product-moment correlation
##
## data:  states20$infant_mort and states20$abortion_laws
## t = 3.9131, df = 48, p-value = 0.0002867
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2472881 0.6774048
## sample estimates:
##       cor
```

```
## 0.4917852
```

The results from Pearson's R show a weak, positive relationship (0.49) between states' infant mortality and number of abortion laws per state. This confirms our initial findings from the scatterplot interpretation, and since Pearson's R results can range from 0 (no relationship to +/-1 (strong positive or negative relationship), we can infer that this relationship is quite weak, especially if compared to the previous analyses. In addition, the p-value for this relationship is 0.0002867, which is very close to zero. Therefore, we can reject the null hypothesis and conclude that the true correlation is not equal to zero.