

Sample Project Report Source

The following is the RMarkdown source code that users would write in order to generate the report. Some of this is boilerplate code that is provided by RStudio.

```
----
title: "Sample Project Report"
author: "Haris Skiadas and Barb Wahl"
date: "11/19/2017"
output: html_document
----

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r, message=FALSE, warning=FALSE, include=FALSE}
library(hanoverbase)
```

## The Iris data set

We will explore the [iris dataset](https://en.wikipedia.org/wiki/Iris_flower_data_set)
that comes with R, based on data collected by Edgar Anderson (1935) and
analyzed by R. A. Fisher (1936).

### Numerical Summaries

We start by loading the data set:
```{r}
data(iris)
summary(iris)
```

We see that the dataset contains four numerical variables and one categorical variable:

1. 'Sepal.Length': The *sepal length*, in centimeters ,
2. 'Sepal.Width': The *sepal width*, in centimeters ,
3. 'Petal.Length': The *petal length*, in centimeters ,
4. 'Petal.Width': The *petal width*, in centimeters ,
5. 'Species': The three different *species* of iris considered:
   [setosa](https://en.wikipedia.org/wiki/Iris_setosa),
   [versicolor](https://en.wikipedia.org/wiki/Iris_versicolor) and
   [virginica](https://en.wikipedia.org/wiki/Iris_virginica)

The 'summary' command gave us five-number summaries for the numerical variables and
a frequency table for the categorical variable.

We can also create these **numerical summaries** on their own:
```{r}
favstats(~Sepal.Length, data=iris)
tally(~Species, data=iris)
```

We can also produce summaries *within each Species*:
```{r}
favstats(~Petal.Length|Species, data=iris)
```

```

...

Basic Graphs

We can easily generate histograms:
```{r}
histogram(~Petal.Length, breaks=20, col="purple", data=iris)
...

```{r, fig.height=7}
histogram(~Petal.Length|Species, data=iris, layout=c(1, 3))
...

And some boxplots:
```{r}
bwplot(Species~Petal.Length, data=iris)
...

A scatterplot:
```{r, warning=FALSE}
xyplot(Sepal.Length~Petal.Length, data=iris, groups=Species,
 fill=brewer.pal(3, "Dark2"), pch=21:23, lwd=2, col="black",
 main="Iris data (green=setosa, orange=versicolor, purple=virginica)",
 xlab="Petal Length (cm)",
 ylab="Sepal Length (cm)",
 type=c("p", "smooth"))
...

A labeled dotplot of the mean Petal Length for each species:
```{r}
mean(~Petal.Length|Species, data=iris) %>% sort() %>% dotplot()
...

### Linear Regression

```{r}
setosaFit <- lm(Sepal.Length~Petal.Length,
 data=iris %>% filter(Species == "setosa"))
summary(setosaFit)
...

As anticipated, the linear model for setosa is weak.

Residual plot:
```{r}
residPlot <- xyplot(resid(setosaFit)~fitted(setosaFit),
                  xlab="Predicted Values", ylab="Residuals")
ladd(panel.abline(h=0, lwd=2, col="black"), plot=residPlot)
...

```