

```
# WEEK 8 - Lecture 1
```

```
# VIDEO 4 - A BASIC SCATTERPLOT
```

```
# Read in data
```

```
WHO = read.csv("WHO.csv")
```

```
str(WHO)
```

```
# Plot from Week 1
```

```
plot(WHO$GNI, WHO$FertilityRate)
```

```
# Let's redo this using ggplot
```

```
# Install and load the ggplot2 library:
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
# Create the ggplot object with the data and the aesthetic mapping:
```

```
scatterplot = ggplot(WHO, aes(x = GNI, y = FertilityRate))
```

```
# Add the geom_point geometry
```

```
scatterplot + geom_point()
```

```
# Make a line graph instead:
```

```
scatterplot + geom_line()
```

```
# Switch back to our points:
```

```
scatterplot + geom_point()
```

```
# Redo the plot with blue triangles instead of circles:
```

```
scatterplot + geom_point(color = "blue", size = 3, shape = 17)
```

```
# Another option:
```

```
scatterplot + geom_point(color = "darkred", size = 3, shape = 8)
```

```
# Add a title to the plot:
```

```
scatterplot + geom_point(colour = "blue", size = 3, shape = 17) +  
  ggtitle("Fertility Rate vs. Gross National Income")
```

```
# Save our plot:
```

```
fertilityGNIplot = scatterplot + geom_point(colour = "blue", size =  
  3, shape = 17) + ggtitle("Fertility Rate vs. Gross National Income")
```

```
pdf("MyPlot.pdf")

print(fertilityGNIplot)

dev.off()
```

VIDEO 5 - MORE ADVANCED SCATTERPLOTS

```
# Color the points by region:
ggplot(WHO, aes(x = GNI, y = FertilityRate, color = Region)) +
  geom_point()

# Color the points according to life expectancy:
ggplot(WHO, aes(x = GNI, y = FertilityRate, color = LifeExpectancy))
+ geom_point()

# Is the fertility rate of a country was a good predictor of the
percentage of the population under 15?
ggplot(WHO, aes(x = FertilityRate, y = Under15)) + geom_point()

# Let's try a log transformation:
ggplot(WHO, aes(x = log(FertilityRate), y = Under15)) + geom_point()

# Simple linear regression model to predict the percentage of the
population under 15, using the log of the fertility rate:
mod = lm(Under15 ~ log(FertilityRate), data = WHO)
summary(mod)

# Add this regression line to our plot:
ggplot(WHO, aes(x = log(FertilityRate), y = Under15)) + geom_point()
+ stat_smooth(method = "lm")

# 99% confidence interval
ggplot(WHO, aes(x = log(FertilityRate), y = Under15)) + geom_point()
+ stat_smooth(method = "lm", level = 0.99)

# No confidence interval in the plot
ggplot(WHO, aes(x = log(FertilityRate), y = Under15)) + geom_point()
+ stat_smooth(method = "lm", se = FALSE)

# Change the color of the regression line:
ggplot(WHO, aes(x = log(FertilityRate), y = Under15)) + geom_point()
+ stat_smooth(method = "lm", colour = "orange")
```

