

What's up with Your Data?

EDP 613

Week 2

Looking at Data Like You're Used to



- We nearly always look at data by rows and columns!
- In Excel, that format is called a cell structure.
- In R, that format is called a data frame.

First Things First!

Let's do the following:

- Create a folder with subfolders by class week (e.g. Week 2, Week 3, etc.) somewhere on your hard drive where you can stick all of the R stuff.
- Go back to RStudio.
- Use **File > New File > R Script** to create a new script.
- Save it with a name ending in `.r` or `.R` in the new folder or subfolder you created.
- Type in `getwd()` in the Console. If that doesn't look right, do this: **Session > Set Working Directory > To Source File Location**

Data Frames: A Very Basic Rundown



- Shows data in rows and columns.
- Is really good for manipulating data.
- Closest thing to Excel.

Get the Gapminder data and load it up

We will work with some of the data from the Gapminder project.

```
install.packages("gapminder")
```

```
library(gapminder)
```

Look at the variables inside a data frame

- To specify a single variable from a data frame, use the dollar sign \$.

```
head(gapminder$lifeExp)
```

```
## [1] 28.801 30.332 31.997 34.020 36.088 38.438
```

```
summary(gapminder$lifeExp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  23.60   48.20   60.71   59.47   70.85   82.60
```

Plotting



- We can visualize the data.

```
hist(gapminder$lifeExp)
```

Getting underlying information about our data



```
class(gapminder$continent)
```

```
## [1] "factor"
```

```
summary(gapminder$continent)
```

```
##   Africa Americas   Asia  Europe Oceania  
##    624      300    396    360     24
```


Levels



```
levels(gapminder$continent)
```

```
## [1] "Africa" "Americas" "Asia" "Europe" "Oceania"
```

- The **levels** of the factor `continent` are "Africa", "Americas", etc.

```
nlevels(gapminder$continent)
```

```
## [1] 5
```

- **nlevels** tells you how many there are.

We can see the data about levels

```
table(gapminder$continent)
```

```
##  
##   Africa Americas   Asia  Europe Oceania  
##     624      300    396    360     24
```

And we can even plot them!



```
barplot(table(gapminder$continent))
```

That's it for today!

