

```
# IRDR04 Research Tools Module
# Week 4 Tutorial
# 23/10/18 9-11am @ Birkbeck, Malet Street 414/415 Public Cluster
# Tutor: Dr Alison Fairbrass, UCL Centre for Biodiversity & Environment Research, alison.fairbrass@gmail.com
```

```
### Walk-through ###
```

```
setwd() # complete with filepath
```

```
# Creating some data
my_score <- c(4,6,9,10,19)
length(my_score)
my_names <- c("Ali", "Jack", "Jim", "Patty", "Izaz")
length(my_names)
df <- cbind(my_score, my_names)
df <- rbind(my_score, my_names)
df <- data.frame(my_score, my_names)
t(df)
```

```
# Normal distribution
# Height
set.seed(10)
height <- rnorm(n=100, mean = 160, sd = 20)
hist(height)
plot(density(height))
```

```
# Summary stats
summary(height)
```

```
# Age
set.seed(10)
age <- rnorm(n=100, mean=60, sd=11)
age <- round(age)
hist(age)
```

```
# Summary stats
summary(age)
```

```
# Create dataframe
df <- data.frame(height, age)
# Look at the column names
names(df)
```

```
# More summary stats
colSums(df)
apply(df, 2, var)
apply(df, 2, sd)
apply(df, 2, mean)
```

```
# Get the means
nrow(df)
colSums(df)/nrow(df)
colMeans(df)
```

```
# Look at the distribution of the data in each column
hist(df$height)
hist(df$age)
```

```
# Use breaks to look in more detail
hist(df$height, breaks = 10)
```

```
# Export and import data
write.csv(df, "ageAndHeight.csv")
dfl <- read.csv("ageAndHeight.csv")
```

```
# Try without indexes
write.csv(df, "ageAndHeight.csv", row.names = F)
dfl <- read.csv("ageAndHeight.csv")
```

```
# Subset the dataset with only person and height data
names(dfl)
df2 <- dfl[c('person', 'height')]
write.csv(df2, "height.csv", row.names=F)
```

```
# pnorm. Given a number or a list it computes the probability that a normally distributed random number will be less than that number.
plot(density(rnorm(10000, mean=22, sd=5)))
```

```
# The probability that a randomly selected normally variable with mean 5 and variance 25 lies below 0
pnorm(10, mean=22, sd=5)
# Or lies below 10
pnorm(35, mean=22, sd=5)
```