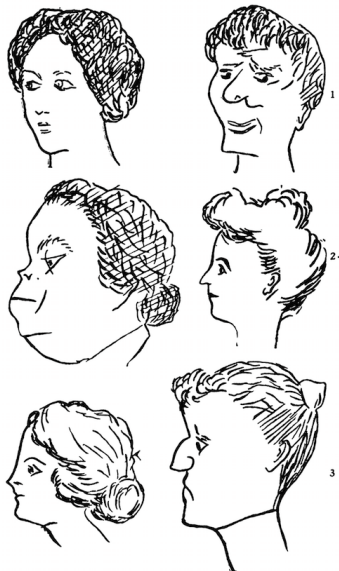


# E-411-PRMA

## Lecture 1

Christopher David Desjardins

17 August 2015



THE PSYCHOLOGICAL CLINIC is indebted for the loan of these cuts and those on p. 225 to the courtesy of Dr. Oliver F. Corman, Associate Superintendent of Schools of Philadelphia, and Chairman of Committee on Backward Children Investigation. See Report of Committee, Dec. 31, 1910, appendix.



*Extreme desespoir.  
Eufferfte Verzweiffung.*



*Colere meslée de Crainte.  
Zorn mit Furcht vermischet.*

SAT®

## ► Topics

- Statistics, Classical Test Theory, Reliability, Validity, Item Response Theory, Generalizability Theory, Equating, and assessments/issues specific to various fields

## ► Assessments

- R computer assignments (30%)
- Item writing activity (5%)
- Midterm exam (25%)
- Final exam (50%)

The screenshot displays the RStudio application window. The top toolbar includes icons for file operations and a search bar. The main editor pane shows an R script with comments and code for data manipulation. The console pane at the bottom left shows the R startup message and help text. The right-hand pane is split into two sections: 'Environment' (showing an empty global environment) and 'Help' (displaying the documentation for the `separate()` function from the `tidyr` package).

```
26_june_intro_R.R x
Source on Save
Run Source

140
141 # view all data
142 precol
143
144 # If you want to see a specific observation, say #20, and you want to see responses from
145 # variables 2 through 3
146 precol[20,2:3]
147
148 # Or if you want to see variables 1, 3, and 6 for observations 19 through 21.
149 precol[19:21, c(1,3,6)]
150
151 # Don't worry R won't be so bad ;)
152
153 # To get information about the data set and the variable types
154 str(precol)
155
156 # To get the variable labels
157 attr(,"names") = "variable.labels"
158 (Top Level) >
```

Console

Copyright (C) 2015 The R Foundation for Statistical Computing  
Platform: x86\_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

>

Environment History

Global Environment

Environment is empty

Files Plots Packages Help Viewer

R: Separate one column into multiple columns. Find in Topic

separate (tidyr) R Documentation

### Separate one column into multiple columns.

#### Description

Given either regular expression or a vector of character positions, `separate()` turns a single character column into multiple columns.

#### Usage

```
separate(data, col, into, sep = "[^a-zA-Z:]+", remove = TRUE,
          convert = FALSE, extra = "error", ...)
```

#### Arguments

<code>data</code>	A data frame.
<code>col</code>	Bare column name.
<code>into</code>	Names of new variables to create as character vector.
<code>sep</code>	Separator between columns.

R: <https://www.r-project.org>  
RStudio: <https://www.rstudio.com>

# Why should I learn R?

- ▶ It's free and open-source

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses



# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it
- ▶ Steep learning curve

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it
- ▶ Steep learning curve
  - ▶ Will provide nearly all the code

# Why should I learn R?

- ▶ It's free and open-source
- ▶ Statistics and psychometrics analyses
- ▶ Helps you learn statistics better
- ▶ Learn reproducible research
- ▶ Extremely marketable skill
- ▶ High quality graphics
- ▶ Everyone is doing it
- ▶ Steep learning curve
  - ▶ Will provide nearly all the code
- ▶ No SPSS in this class



# Resources for R

- ▶ Icelandic resources

  - <http://kennslubanki.hi.is/search/efni/r>

  - <http://kennslubanki.hi.is/tolfraedi/myndbond/rrstudio-inngangur>

  - <http://kennslubanki.hi.is/tolfraedi/myndbond/rrstudio-fyrstu-skrefin>

- ▶ Please watch the last two videos before next class
- ▶ Please install R and RStudio before next class
- ▶ Next class will be an R workshop

What is **measurement**?

What is **measurement**?

What is a **test**?

What is **measurement**?

What is a **test**?

What is a **scale**?

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

What is a **scale**?

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

An instrument used to measure

What is a **scale**?

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

An instrument used to measure

What is a **scale**?

A set of numbers used to categorize or quantify variables  
("things")

What is **measurement**?

Assignment of numerical values based on a set of rules

What is a **test**?

An instrument used to measure

What is a **scale**?

A set of numbers used to categorize or quantify variables  
("things")

**Nominal**

**Ordinal**

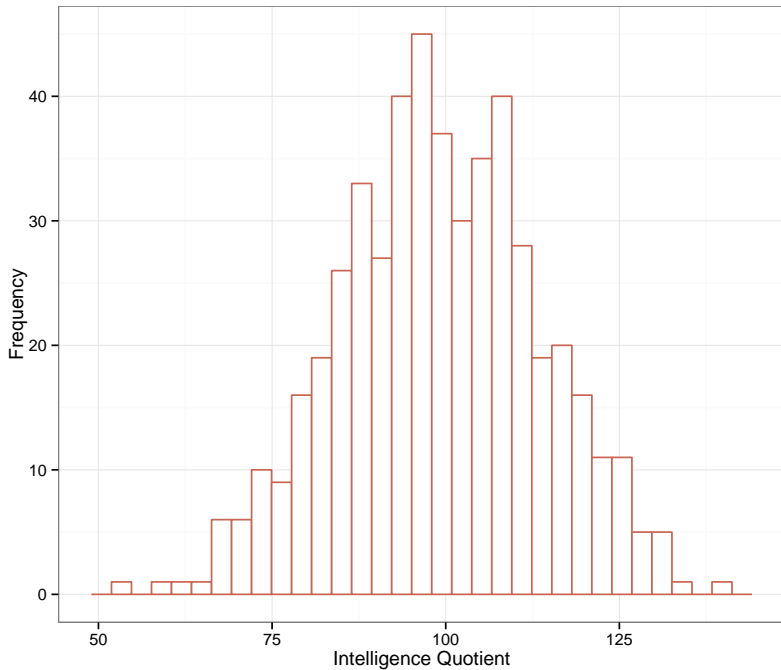
**Ratio**

**Interval**



# What kind of scales are these?

- ▶ Temperature
- ▶ Height
- ▶ Intelligence Quotient
- ▶ Color
- ▶ Ethnic group
- ▶ Likert-type items
- ▶ Job satisfaction

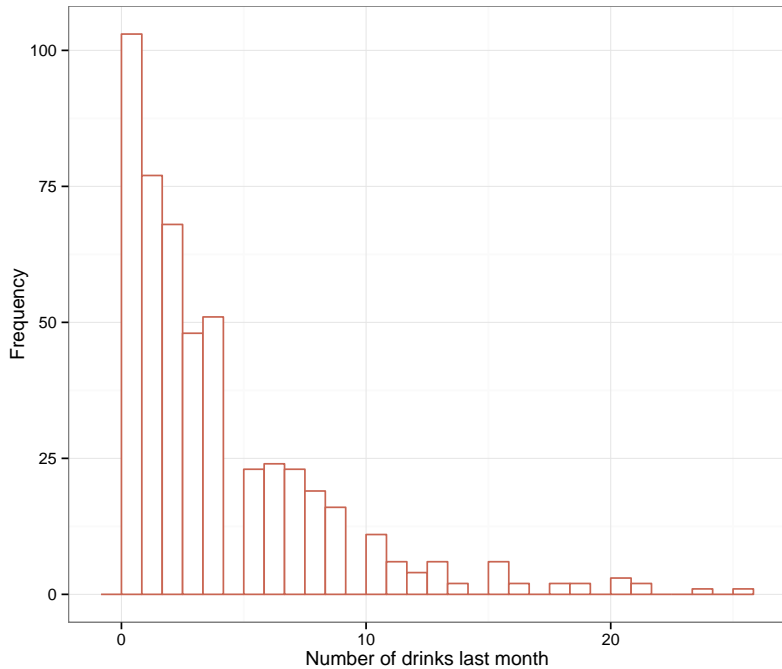


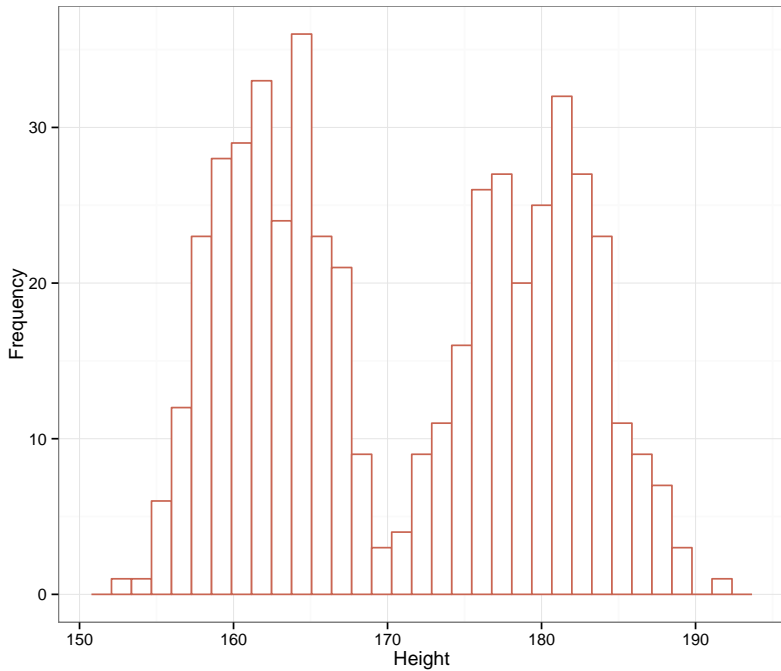
```
# Load the library
set.seed(101)
library("ggplot2")

# Set up the parameters
sample_size <- 500
mean <- 100
standard_deviation <- 15

# Generate random numbers
x <- rnorm(sample_size, mean, standard_deviation)

# Plot the data
qplot(x, fill = I("white"), color = I("#c96552")) +
  theme_bw() + xlab("Intelligence Quotient") +
  ylab("Frequency")
```





# Central Tendency Measures

## Mean

$$\bar{X} = \frac{\sum X_i}{n}$$

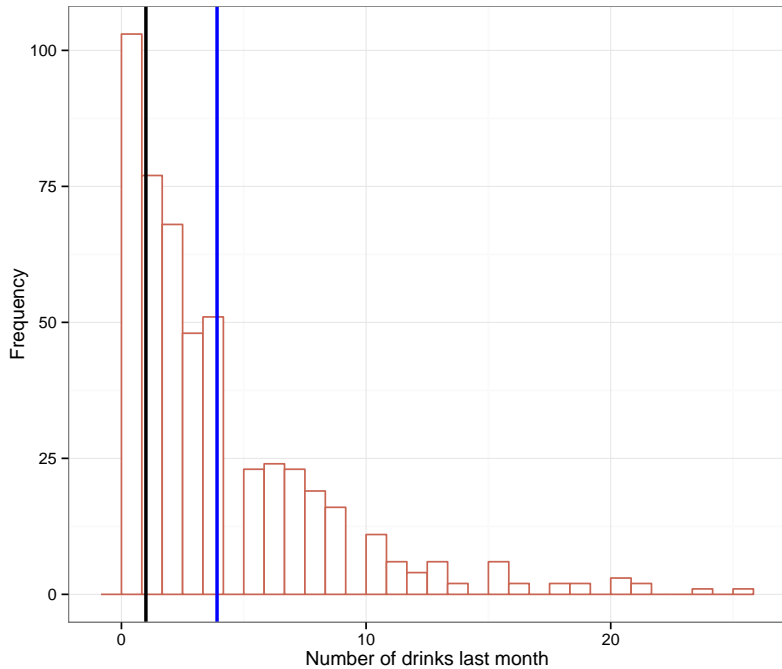
## Median

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2}$$

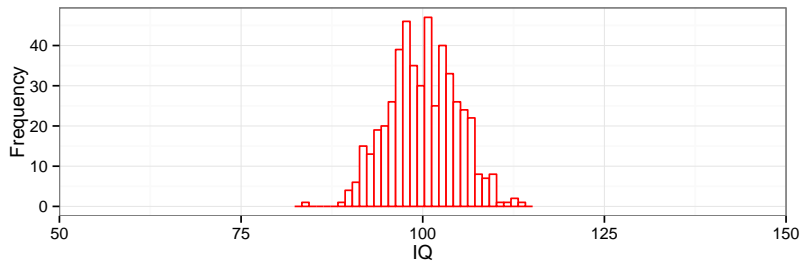
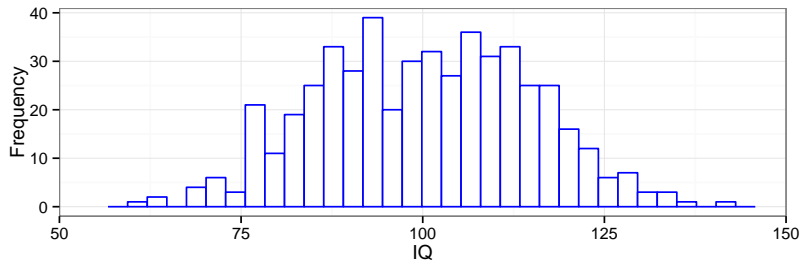
## Mode

The most frequently occurring value

Which of these statistics is most robust to outliers?



# Variability





# Measures of variability

Range

Interquartile range ( $Q_1$ ,  $Q_2$ ,  $Q_3$ )

Standard Deviation and Variance

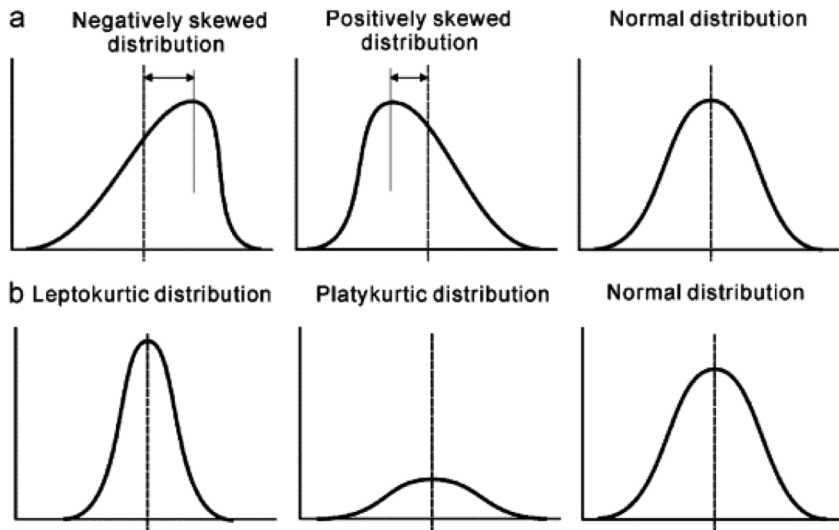
$$s = \sqrt{\frac{\sum X_i - \bar{X}}{n - 1}}$$

$$s^2 = \frac{\sum X_i - \bar{X}}{n - 1}$$

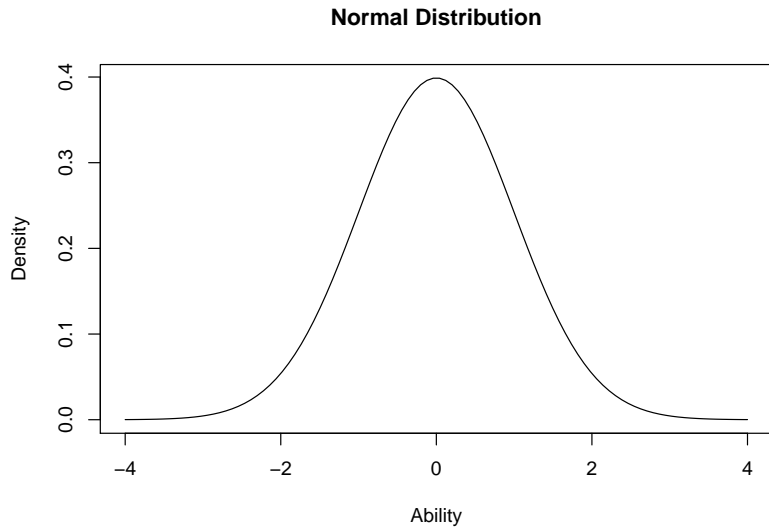
# Distributions, skewness, kurtosis

- ▶ What is a probability distribution
  - ▶ Assigns a probability, likeliness of occurrence, of a score of all possible scores
  - ▶ May be parametric or non-parametric
- ▶ What skew might you expect these outcomes to look like?
  - ▶ Reaction time in a psychological experiment
  - ▶ Number of children in a family
  - ▶ Scores on an easy test
  - ▶ Height in Iceland
- ▶ Platykurtic, mesokurtic, and leptokurtic
- ▶ Plot your data, rely less on statistics!

# Shapes of distributions



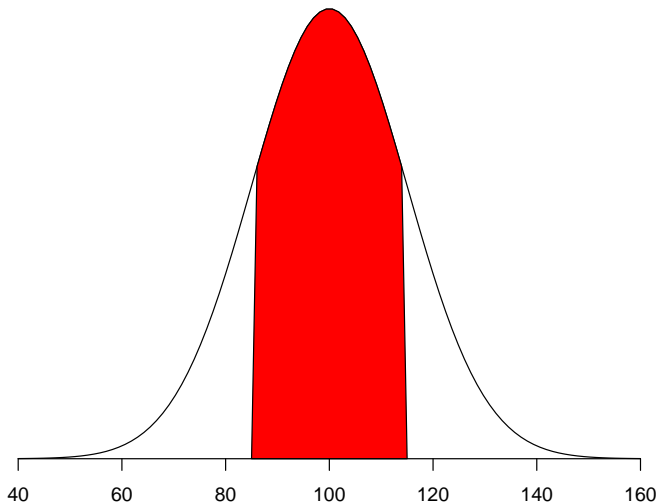
# Normal Distribution



# IQ - 1 Standard Deviation

## Normal Distribution

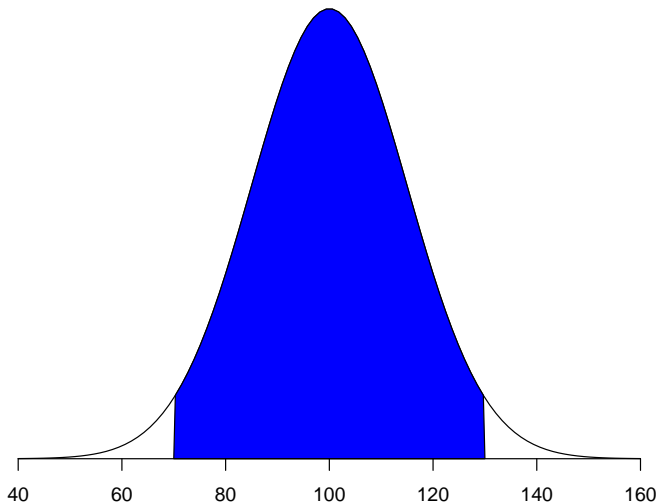
$$P(85 < IQ < 115) = 0.683$$



# IQ - 2 Standard Deviation

## Normal Distribution

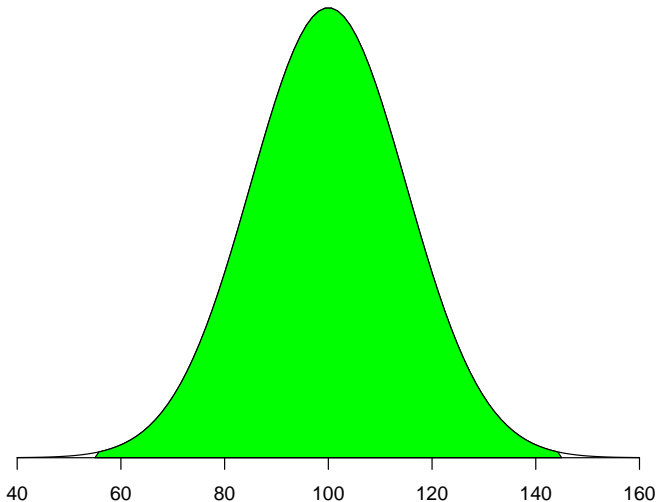
$$P(70 < IQ < 130) = 0.954$$



# IQ - 3 Standard Deviation

## Normal Distribution

$$P(55 < IQ < 145) = 0.997$$



# Characteristics of the Normal distribution

