

# CHI Workshop 2023

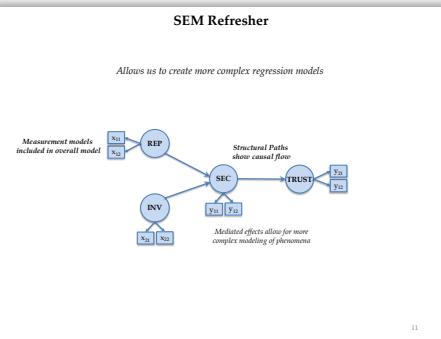
## SEMinR

### Getting Started SEM, R, and SEMinR

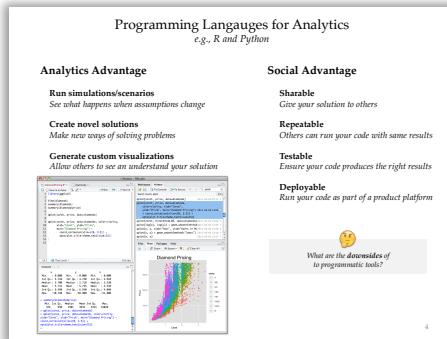
### Setting Up Your Model Integrating SEM into HCI Research

### Evaluation and Reporting Methods, Tables, and Figures

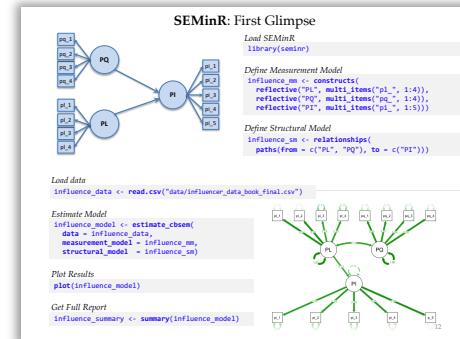
#### About SEM



#### Programming in R



#### SEMinR: First Glimpse

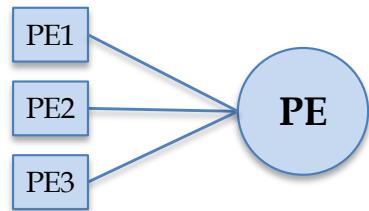


# **Introduction to SEM**

*(Structural Equation Modeling)*

# SEM Example: *Technology Adoption Survey*

## Measurement Items



*Abstract concepts are statistically constructed of multiple items*

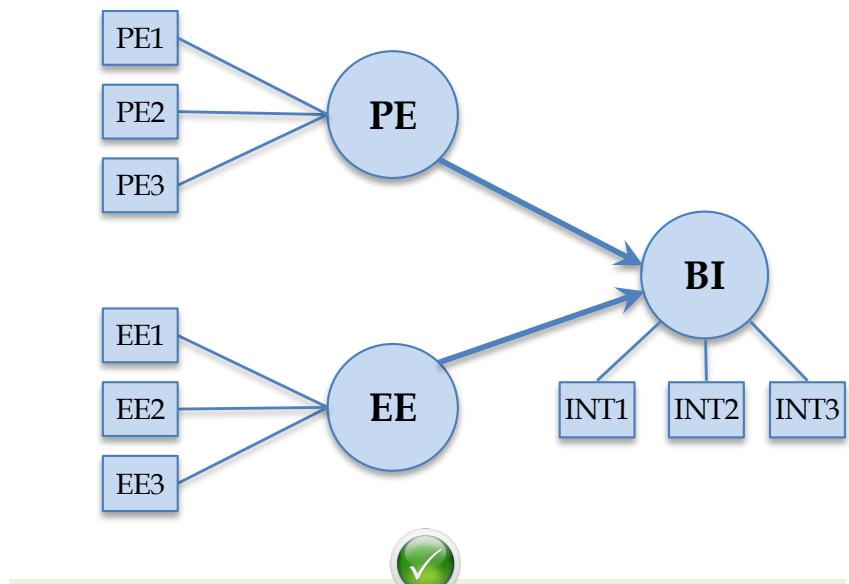
## Measurement Items

### PE: Performance Expectancy

1. I would find Trello useful in my social, work, or personal projects.
2. Using Trello would enable me to accomplish tasks more quickly.
3. Using Trello would increase my productivity.

## Structural Model

*Constructs can be related to one another in a regression-type model*



*Relating constructs is more parsimonious than relating all the measurement items*

### EE: Effort Expectancy

1. Learning to operate Trello would be easy for me.
2. I would find it easy to get Trello to do what I want it to do.
3. My interaction with Trello would be clear and understandable.
4. It would be easy for me to become skillful at using Trello.

### BI: Behavioral Intention

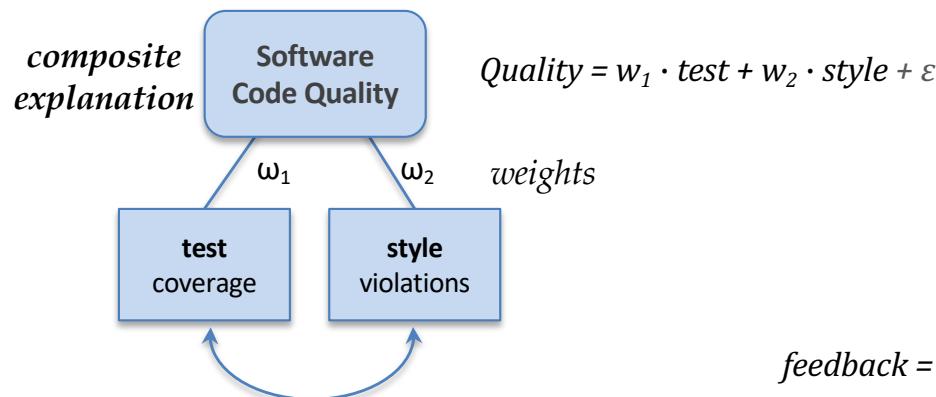
1. I intend to use Trello in the next month.
2. I predict I would use Trello in the next month.
3. I plan to use Trello in the next month.

# Measurement Models

Measuring Phenomena from Observed Variables

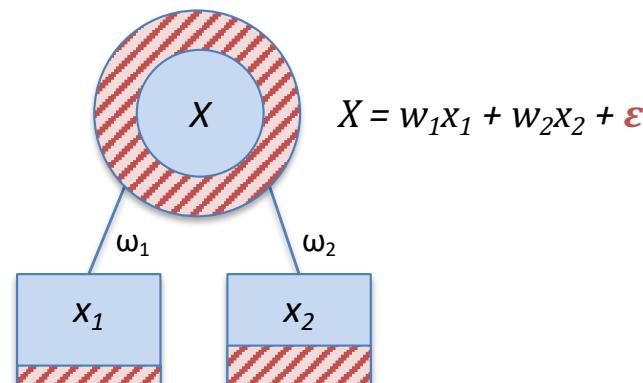
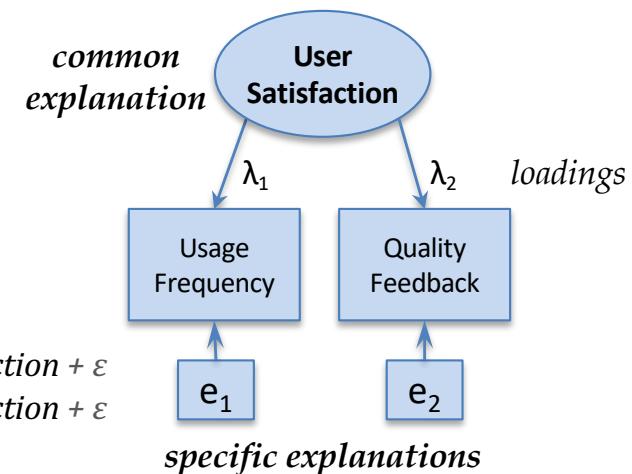
## Composite Model

Composite constructs are aggregates of items  
(useful for prediction/recreating data)

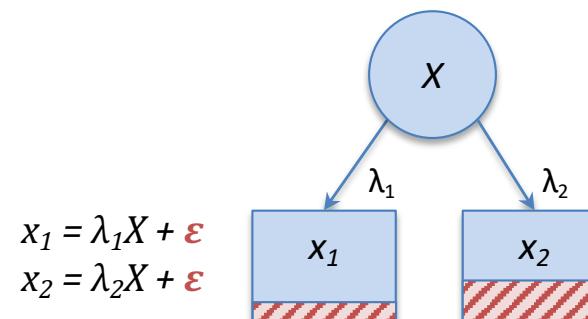


## Reflective Factor Model

Latent constructs are free of measurement error!  
(useful for explanation/interpretation)



composite measurement combines shared variance, *unique variance*



Reflective Measurement captures only the common variance of items, and not its *specific variances* (measurement errors?)

# Structural model

Estimating the Relationship Between Constructs

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

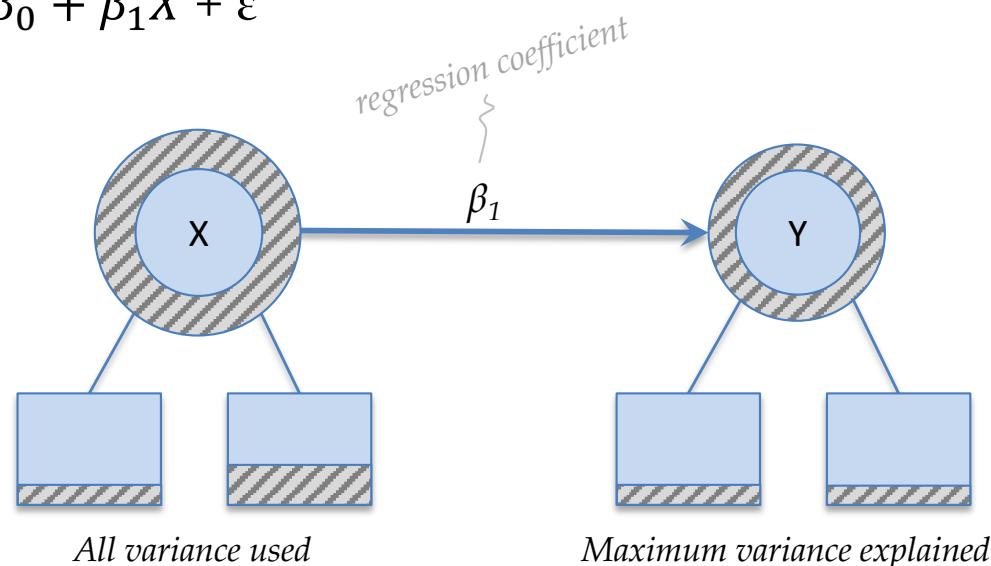
## Composite Models

### Suitable for Explanation

*Correct way to model known composites*

### Better for Prediction

*Composite model maximizes captured variance  
Better for predicting actual outcomes*



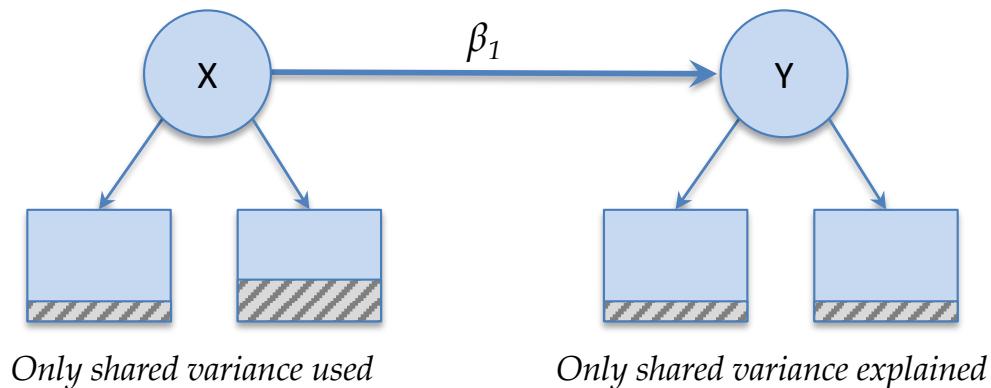
## Factor Models

### Better for Explanation

*Factor model maximizes shared variance  
Useful for modeling pure latent, abstract factors*

### NOT suitable for prediction

*Factors scores are not determinable*

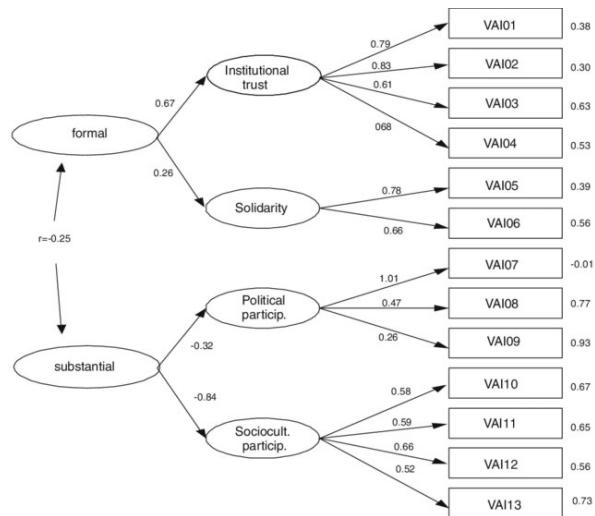


# Structural Equation Models (SEM)

Using constructs to model concepts measured by multiple items

## Social Sciences

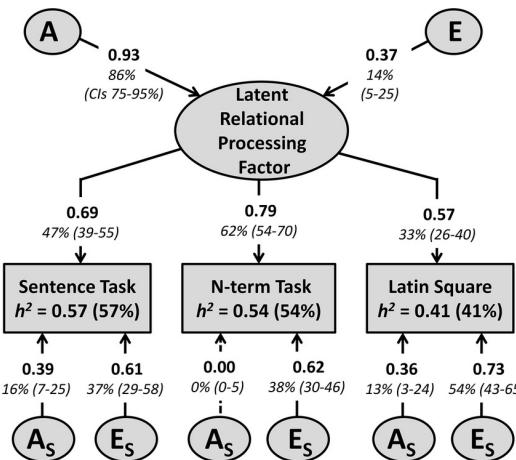
(Psychology/Sociology/Management)



[https://www.researchgate.net/publication/257664169\\_Construction\\_Validation\\_and\\_Application\\_of\\_the\\_Measurement\\_of\\_Social\\_Cohesion\\_in\\_47\\_European\\_Countries\\_and\\_Regions/figures?lo=1](https://www.researchgate.net/publication/257664169_Construction_Validation_and_Application_of_the_Measurement_of_Social_Cohesion_in_47_European_Countries_and_Regions/figures?lo=1)

## Life Sciences

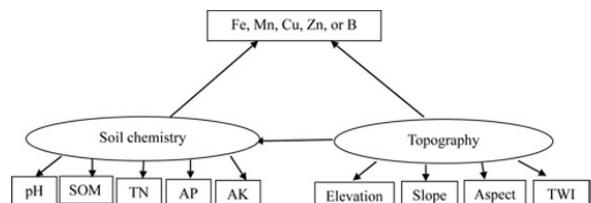
(Genetics/Neuroscience)



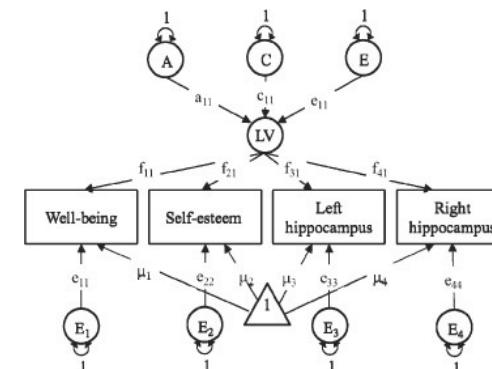
[https://openi.nlm.nih.gov/detailedresult.php?img=PMC4393228\\_pone.0123886.g002&req=4](https://openi.nlm.nih.gov/detailedresult.php?img=PMC4393228_pone.0123886.g002&req=4)

## Natural Sciences

(Chemistry)



<http://www.scielo.cl/scielo.php?pid=S0718-95162016005000076>



[https://www.researchgate.net/publication/223984074\\_A\\_multivariate\\_twin\\_study\\_of\\_hippocampal\\_volume\\_self-esteem\\_and\\_well-being\\_in\\_middle-aged\\_men](https://www.researchgate.net/publication/223984074_A_multivariate_twin_study_of_hippocampal_volume_self-esteem_and_well-being_in_middle-aged_men)

# **Introduction to R + RStudio**

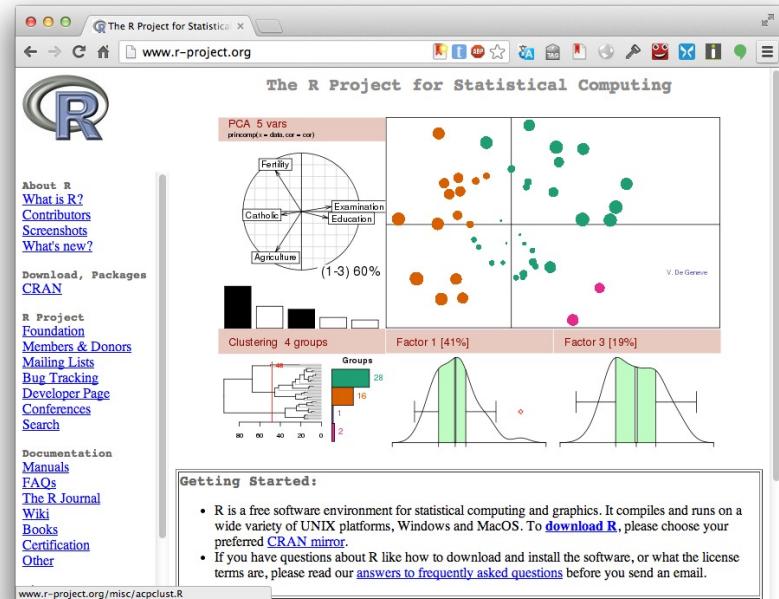
*(Statistical Programming Language + Development Environment)*

# R and Rstudio on your Computer

Install R

<http://www.r-project.org/>

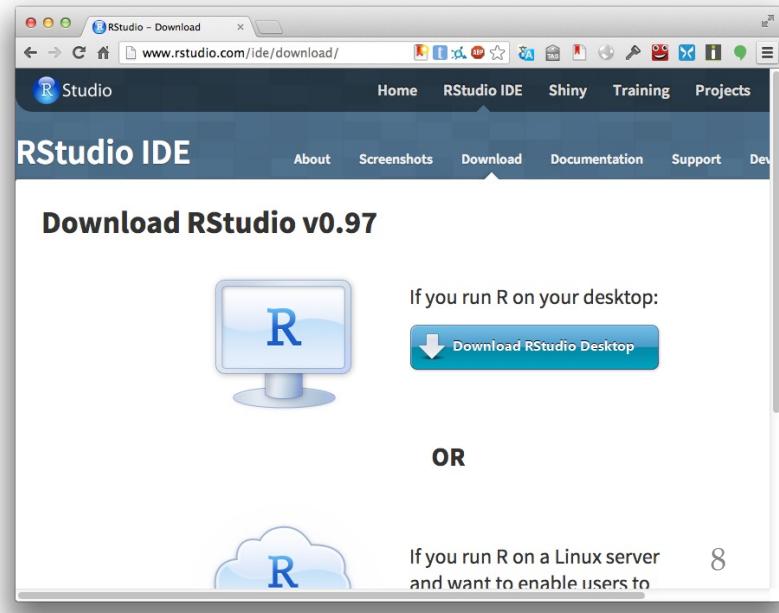
*Programming language platform*



Install RStudio

<http://www.rstudio.com/>

*Integrated Development Environment*



# posit.cloud: R and RStudio in the Cloud

“Spaces”  
(collections of projects)

The screenshot shows the posit.cloud web interface. On the left, there's a sidebar with navigation links: Spaces (Your Workspace, compstatslib issues, SEMinR Course CHI2023, New Space), Learn (Guide, What's New, Primers, Cheat Sheets), and Help (Current System Status, Posit Community). The main content area is titled "SEMinR Course CHI2023" and shows a list of "All Content" (4 items). Each item has a title, a brief description, and creation details. A red oval highlights the "New Project" button in the top right of the content area. The URL in the browser bar is https://posit.cloud/spaces/362061/content/all?sort=name\_asc.

SEMinR Course CHI2023

Content Data Members Usage About

New Project

All Content (4)

TYPE \* ACCESS \* SORT AZ

Before the course starts

Welcome to posit.cloud. This project is only a demo project. The course sessions will unlock as soon as we get there :)

R RStudio Project André Calero Valdez Space members Created Apr 12, 2023 5:50 PM

Session 1 - Getting started

This is the project for the first session.

R RStudio Project André Calero Valdez Private Created Apr 11, 2023 9:13 PM

Session 2 - Setting up your model

This is the project for the second session.

R RStudio Project André Calero Valdez Private Created Apr 11, 2023 9:14 PM

Session 3 - Evaluating & reporting your model

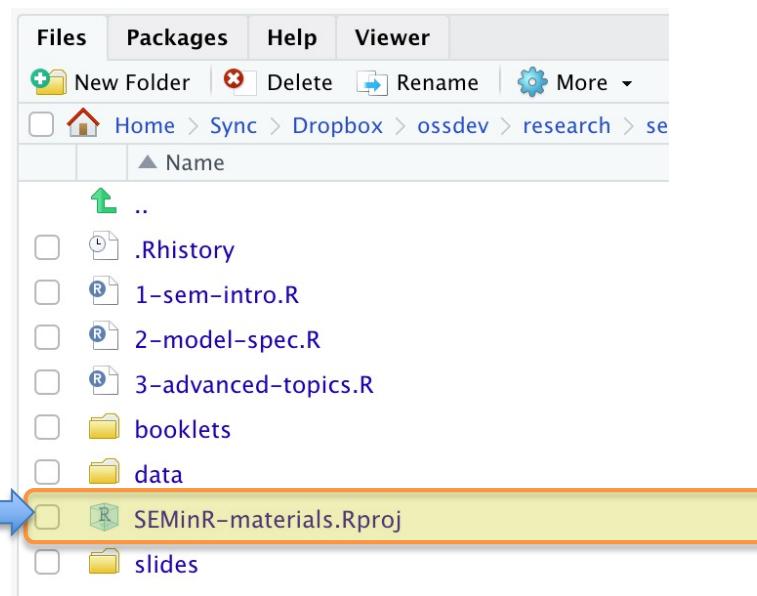
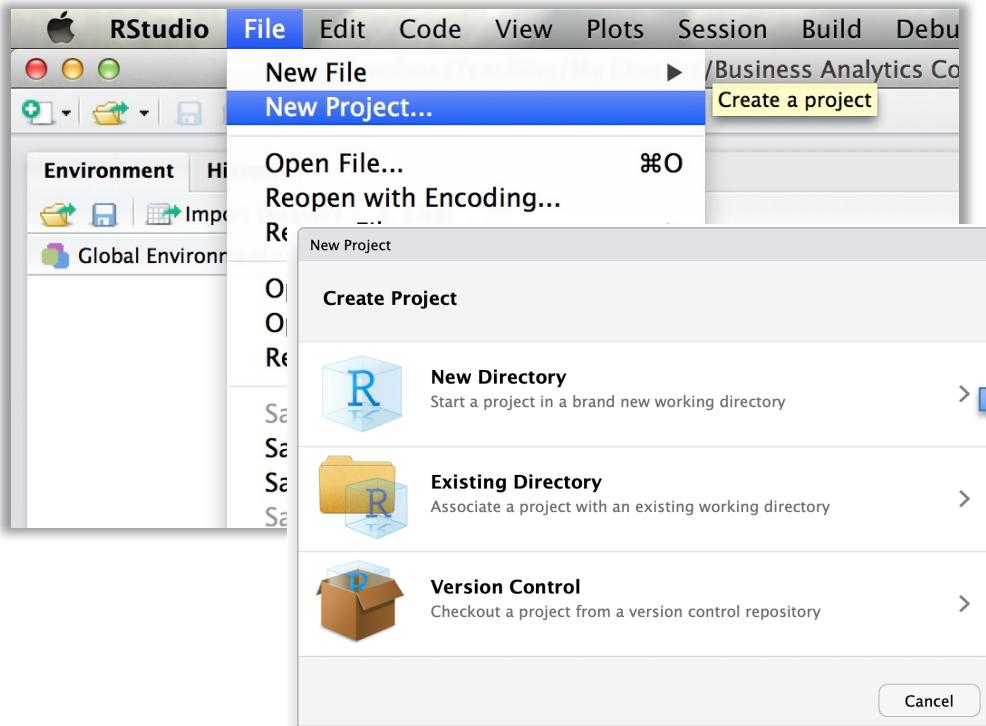
This is the project for the third session.

R RStudio Project André Calero Valdez Private Created Apr 11, 2023 9:21 PM

https://posit.cloud/spaces/362061/content/5771798

“Projects”  
(collections of code/data)

# RStudio: Using Projects



Project metadata is stored in an **Rproj** file  
*This file does NOT contain your code/data*

**Make a new RStudio project** for each of your:

- Research projects
- Assignments
- Demonstrations
- Etc.

# RStudio: Workspace

**Folders and Files**

SEMinR-materials - RStudio

SEMinR-materials

Files Packages Help Viewer

New Folder Delete Rename More

research > sem-in-r > SEMinR Development > workshops > NWU SEMinR Workshop > SEMinR-materials

Name Size Modified

.. .Rhistory 0 B May 10, 2021, 9:56 PM  
1-sem-intro.R 6.5 KB May 10, 2021, 9:52 PM  
2-model-spec.R 9.2 KB May 10, 2021, 9:42 PM  
3-advanced-topics.R 7.4 KB May 10, 2021, 7:56 PM  
booklet  
data  
SEMinR  
slides

**Plots**

PI

PL

**Scripting**

2-model-spec.R

Source on Save Run Source

```
152 # using covariance-based SEM – we one again leverage t.  
153 # package to estimate for CBSEM.  
154  
155 influence_sm <- relationships(  
156   paths(from = c("PL", "SIC"), to = "PI"))  
157  
158 cbsem_model <- sem(data = user_data, influen  
159  
160 plot(cbsem_mo [c("PL", "PI")])  
161  
162 # However, a full report of results comes again from t.  
163  
164 cbsem_report <- summary(cbsem_model)  
165  
166
```

162:1 (Top Level) R Script

**R Console**

Console Terminal Jobs

~/Sync/Dropbox/ossdev/research/sem-in-r/SEMinR Development/workshops/NWU SEMinR Workshop/SEMinR> plot(cbsem\_model\$construct\_scores[, c("PL", "PI")])>

# Learning R



01: Basic Building Blocks

02: Workspace and Files

03: Sequences of Numbers

04: Vectors

06: Subsetting Vectors

13: Simulation

09: Functions

10: lapply and sapply

05: Missing Values

08: Logic

14: Dates and Times

# R: Basic Commands

## Variables

```
value <- 5  
  
construct_name <- "PE"
```

## Collections of values

```
numbers <- c(1, 2, 3, 4, 5)  
  
numbers <- 1:5  
  
item_names <- c("PE1", "PE2")  
  
construct_definition <- list(  
  name = "Performance Expectancy",  
  items = c("PE1", "PE2", "PE3", "PE4"))
```

**Vector:** A sequence of data elements of the same type  
1:5 is a shortcut to create a numeric sequence

**Lists:** collections of values, vectors, or even other lists

## Extracting values from collections

```
item_names[2]  
[1] "PE2"  
  
construct_definition$items  
[1] "PE1" "PE2" "PE3" "PE4"
```

Extract elements from a vector using its [index] position

Extract elements from a list using the \$ notation

## Inspecting objects

See the class (type) of object

```
class(construct_definition)  
"list"
```

See the structure of an object

```
str(construct_definition)  
List of 2  
 $ name : chr "Performance Expectancy"  
 $ items: chr [1:4] "PE1" "PE2" "PE3" "PE4"
```

# Operations

## Math

How to use two inputs to produce an output

$$y = 5 + 2 \quad \text{Addition operation}$$

$$y = 5^2 \quad \text{Exponent operation}$$

## R Programming

Addition

```
5 + 2  
[1] 7
```

Assignment    Sequence

```
numbers <- 1:5  
[1] 1 2 3 4 5
```

Exponent

```
numbers^2  
[1] 1 4 9 16 25
```

# Functions

## Math

How variable  $y$  is associated with variable  $x$

$$y = f(x_1, x_2)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

## R Programming

Function name

```
seq(1, 5)  
[1] 1 2 3 4 5
```

```
sum(numbers)  
[1] 15
```

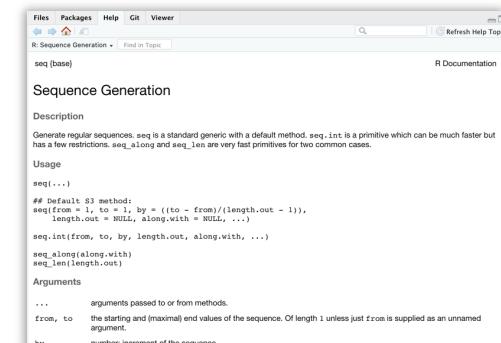
Return Value (Output)

## Getting Help

`help(seq)`

or

`?seq`



Note: R can also behave like an object-oriented programming language...

# Data Analysis Workflow: Linear Regression

## Loading data

```
course_data <- read.csv("course_data.csv")
```

## Specifying + Estimating Model

```
intention_regr <- lm(formula = INT1 ~ PE1 + EE1, data = course_data)
```

## Reporting

```
summary(intention_regr)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.88713   0.57586   1.541   0.125
PE1          0.71345   0.06362  11.214  <2e-16 ***
EE1         -0.14753   0.09372  -1.574   0.117

regr_report <- summary(intention_regr)

str(regr_report)
List of 11
 $ call      : Language lm(formula = INT1 ~ PE1 + EE1, data = course_data)
 $ residuals : Named num [1:216] 1.151 -0.569 0.57 -1.569 -0.283 ...
 $ r.squared : num 0.374
 ...
 
regr_report$coefficients
[1] 0.3744886
```

# **Introduction to SEMinR**

*(SEM Package for R)*

# Installing SEMinR

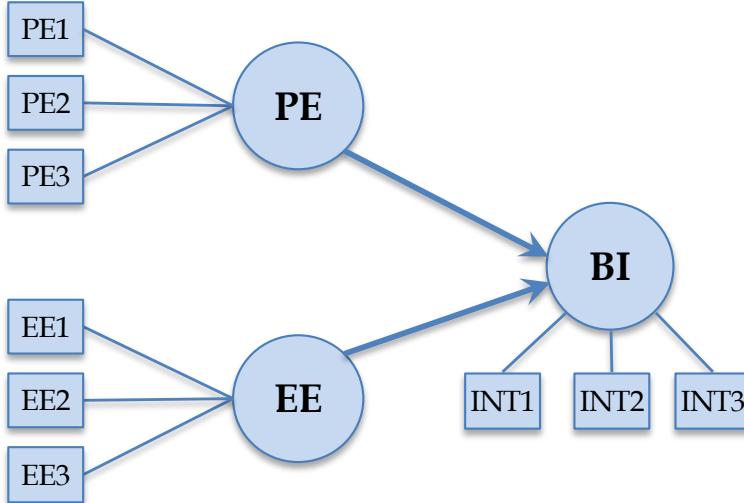
Type the following function call in your RStudio console:

```
install.packages("seminr", dependencies = TRUE)
```

That's it -- you're done installing!



# SEMinR: First Glimpse



Load SEMinR (once everytime your restart RStudio)  
library(seminr)

## Define Measurement Model

```
intention_mm <- constructs(
  reflective("PE", multi_items("PE", 1:3)),
  reflective("EE", multi_items("EE", 1:4)),
  reflective("BI", multi_items("INT", 1:3))
)
```

## Define Structural Model

```
intention_sm <- relationships(
  paths(from = c("PE", "EE"), to = "BI")
)
```

## Load data

```
course_data <- read.csv("course_data.csv")
```

## Estimate Model

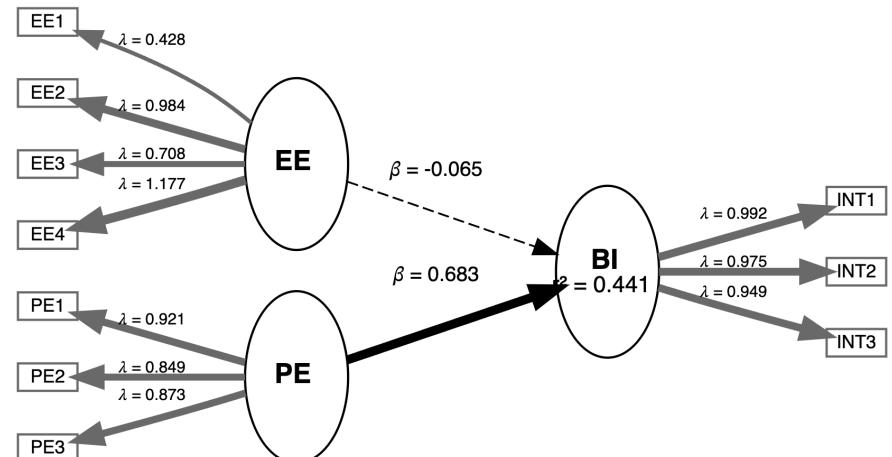
```
intention_model <- estimate_pls(
  measurement_model = intention_mm,
  structural_model = intention_sm,
  data = course_data)
```

## Get Full Report

```
summary(intention_model)
```

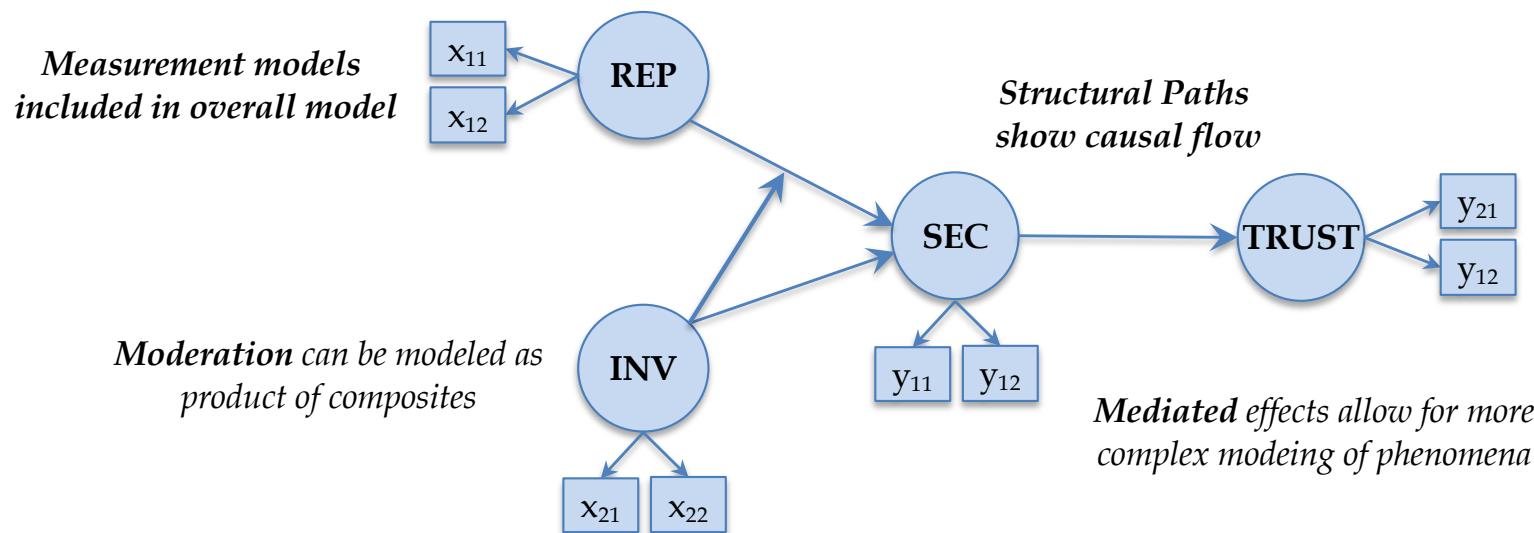
## Visualize Results

```
plot(intention_model)
```



# Measurement, Multiple Outcomes, Moderation, Mediation

*Allows us to create more complex regression models*



# The SEMinR Community

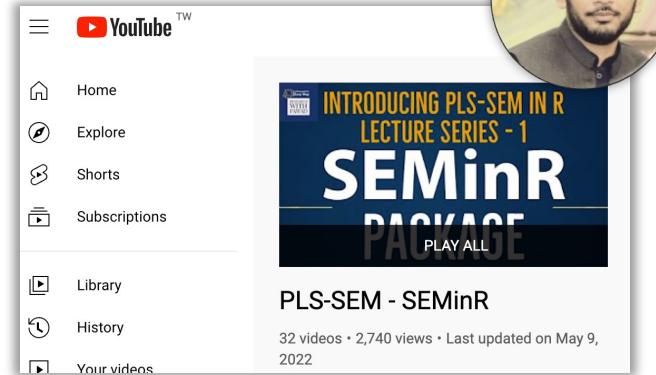


*Writing open-source code is building a community!*

*Others contribute by:*

- Helping other users*
- Bug reporting / fixes*
- Documentation (books, videos)*
- Spreading the word!*

## Video Series



YouTube TW

- Home
- Explore
- Shorts
- Subscriptions
- Library
- History
- Your videos

**INTRODUCING PLS-SEM IN R  
LECTURE SERIES - 1  
SEMinR PACKAGE**

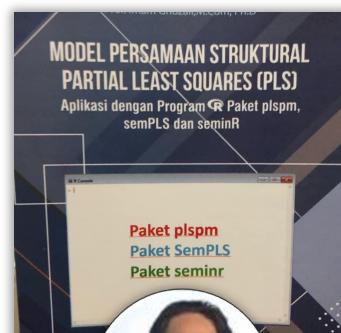
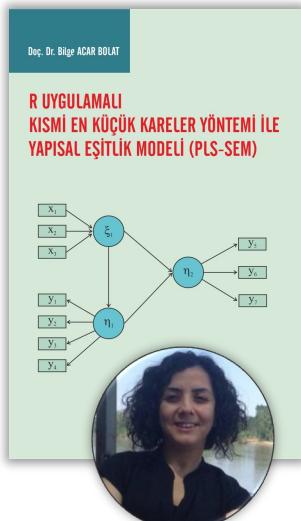
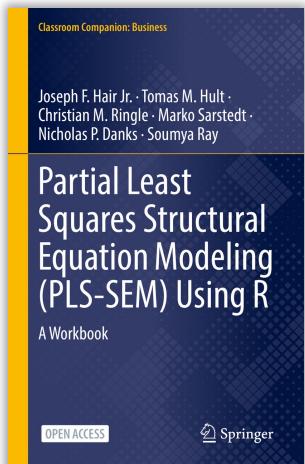
PLAY ALL

PLS-SEM - SEMinR

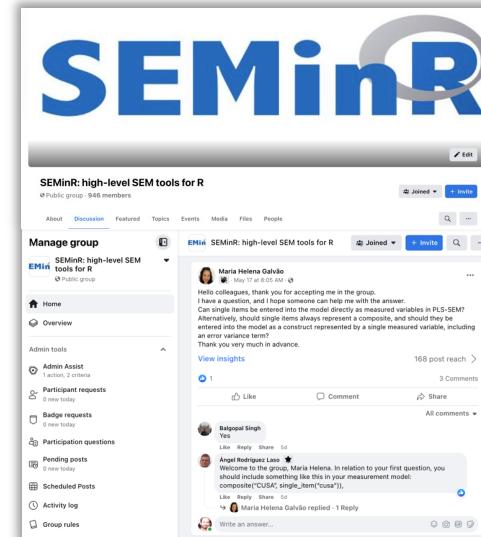
32 videos • 2,740 views • Last updated on May 9, 2022

[https://www.youtube.com/playlist?list=PLb7vm6tsQ3Ks0TyMlw3EUlg0Mr06\\_7U8S](https://www.youtube.com/playlist?list=PLb7vm6tsQ3Ks0TyMlw3EUlg0Mr06_7U8S)

## Published Books



## Facebook Group



**SEMinR**

SEMinR: high-level SEM tools for R

Discussion

168 post reach

168 comments

168 post reach

168 comments

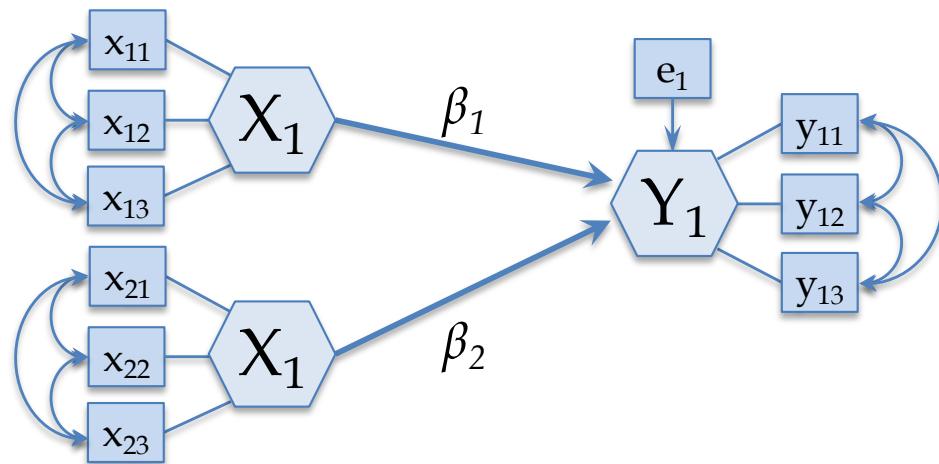
<https://www.facebook.com/groups/seminr>

# **Supplements**

# SEM with Composites

*Partial Least Squares – Path Modeling (PLS-PM)*

**Composite Measurement**  
*(but factors can be estimated)*



## Non-parametric:

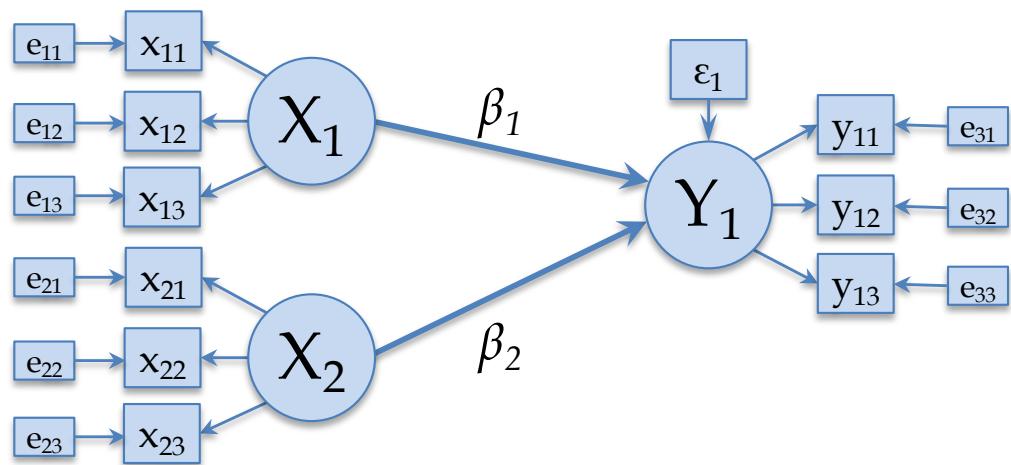
- no global solution formula
- no global goodness-of-fit
- no inferential statistics (standard errors, etc.)

*“more oriented to optimizing explained variances”*

# SEM with Common Factors

*Covariance-based SEM (CB-SEM)*

## Common Factor Measurement (factor scores are indeterminate)



### Parametric:

- Global solution and goodness-of-fit
- Inferential statistics (standard errors, etc.)

*“oriented towards modeling latent constructs”*

# SEM Software

## R packages

PLS-PM

**sem-pls**

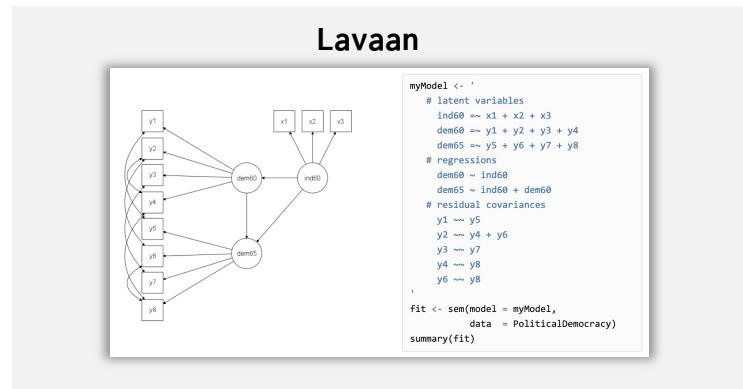
```
R> data("ECSIm")
R> ECSIm
```

**plsPM**

```
# path matrix (inner model realtionships)
ACRIN = c(0, 0, 0)
INDEV = c(0, 0, 0)
POLINS = c(1, 1, 0)
rus_path = rbind(ACRIN, INDEV, POLINS)
```

```
source      target
[1,] "Image"    "Expectation"
[2,] "Expectation" "Quality"
[3,] "Expectation" "Value"
[4,] "Quality"   "Value"
[5,] "Image"    "Satisfactio
[6,] "Expectation" "Satisfactio
[7,] "Quality"   "Satisfactio
[8,] "Value"     "Satisfactio
[9,] "Satisfaction" "Complaints"
[10,] "Image"   "Loyalty"
[11,] "Satisfaction" "Loyalty"
[12,] "Complaints" "Loyalty"
```

CB-SEM



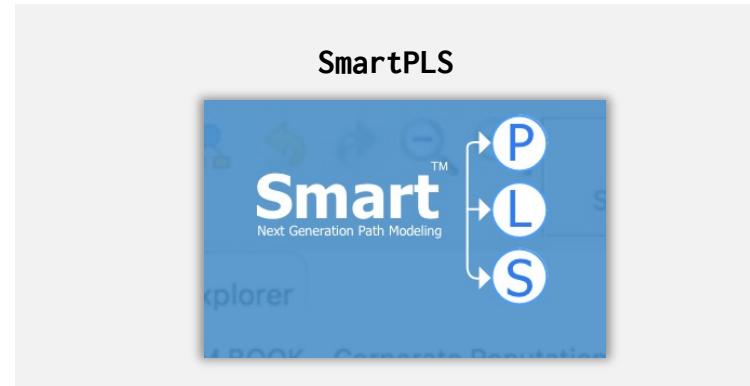
### Advantages

*Open source - extendable with code  
Free!*

### Disadvantages

*Hard to write syntax*

## Commercial Software



### Advantages:

*Very easy to use!*

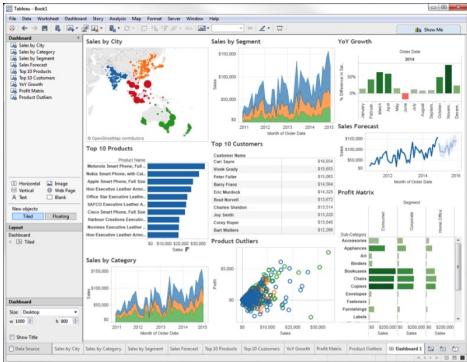
### Disadvantages

*Not free*

*Cannot extend their features*

# Traditional Business Analytics Tools

## Graphical User Interface



Easy to Learn and Use

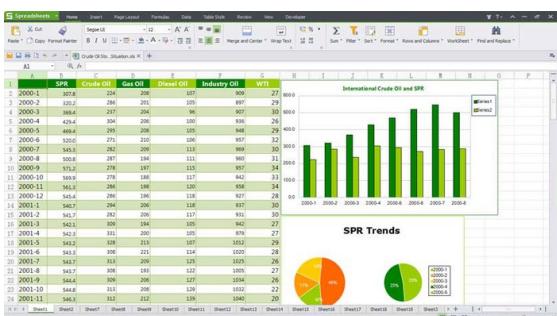
Easy to Analyze Data

Easy to Visualize and Communicate



What are the *downsides* of being limited to GUI and spreadsheet tools?

## Spreadsheets



Familiar Metaphor (Balance Books)

Easy to Manipulate Data

Quick Results

# Computational Tools

## Analytics Advantage

Run simulations/scenarios

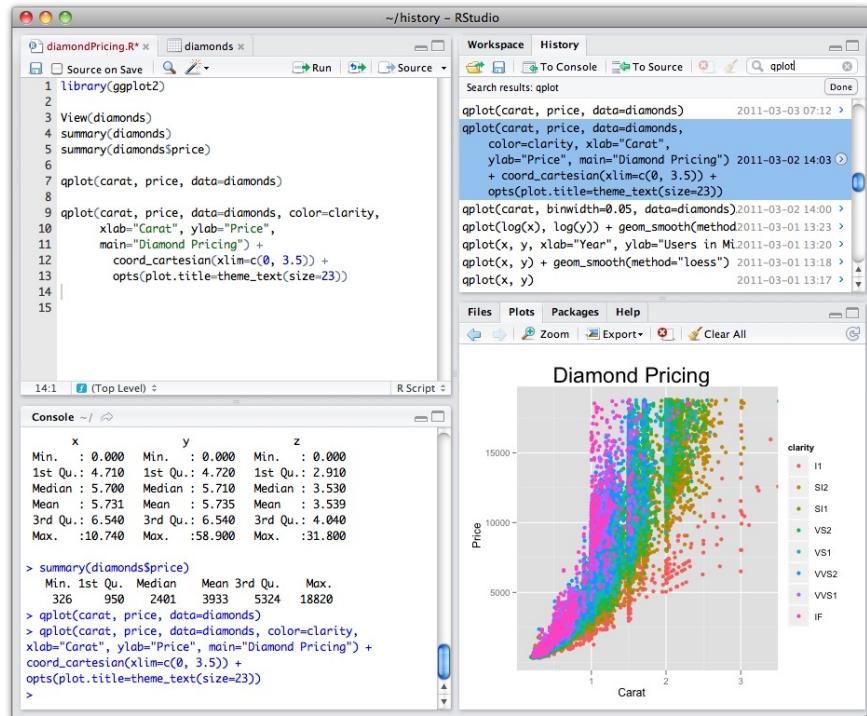
*See what happens when assumptions change*

Create novel solutions

*Make new ways of solving problems*

Generate custom visualizations

*Allow others to see and understand your solution*



*What are the **downsides** of being limited  
to programmatic tools?*

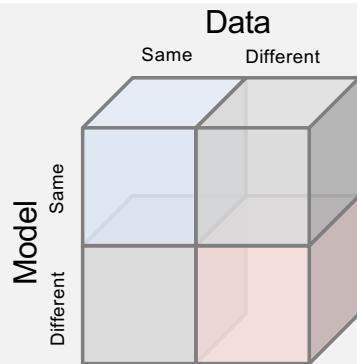
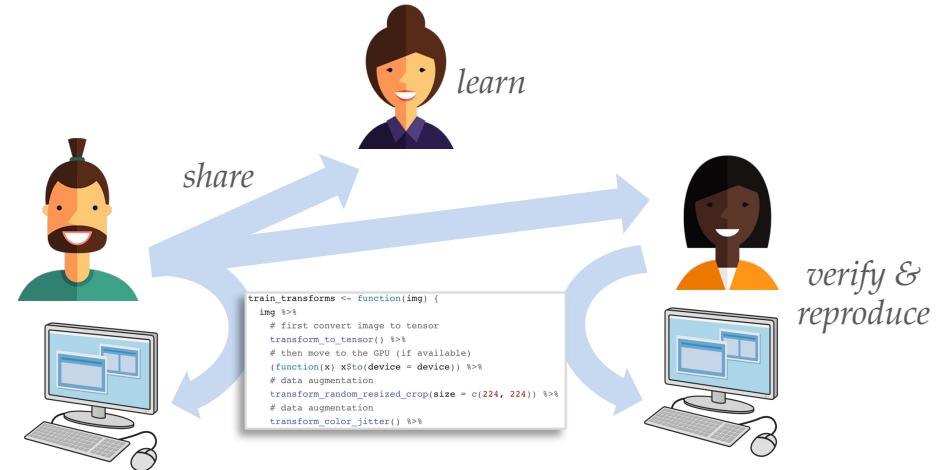
# Coding is Social

## Sharable

*Give your solution to others to learn from*

## Verifiable

*Others can ensure it does the right things*



## Reproducible

*Others can recreate your results*

## Reappliable

*Others can reapply your process*

## Publishable

*Publish your code & data as part of your paper*

## Extensible

*Alter and improve on other studies!*

