



Using RStudio.Cloud to advance R proficiency: a crowdsourcing training experience

R/Pharma 2019

Paulo R. Bargo, Rudradev Sengupta, Bill Pikounis and Surya Mohanty
Statistics & Decision Sciences, Janssen R&D

Rhonda Fenwick, *Time is Now I*

Through her art, Rhonda has explored psoriasis, a chronic skin disorder she has lived with since the age of six.



Outline

- The challenge
- Crowdsource training
 - Setting up
 - Distributing workload
 - Assignment workflow
- RStudio.Cloud
- Resources
- Curriculum
- Learnings

The challenge

- Increase interest in using R for drug submissions
- Very heterogeneous R proficiency in workforce
- Difficult to set aside time for training
- Difficult to retain knowledge with standard methods

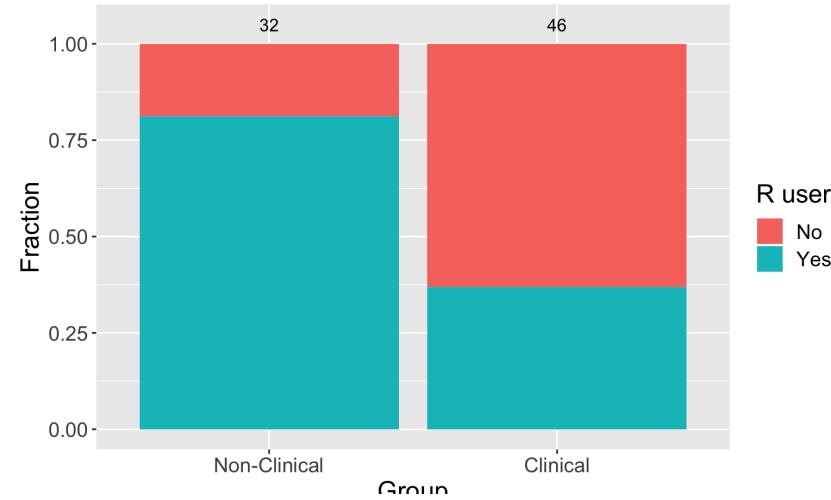


Table 2.1 Amount of Learning vs. Instructional Method (source: Association for Supervision and Curriculum Development Guide, 1988 as cited in Biggs, 1999)

Most People Learn...

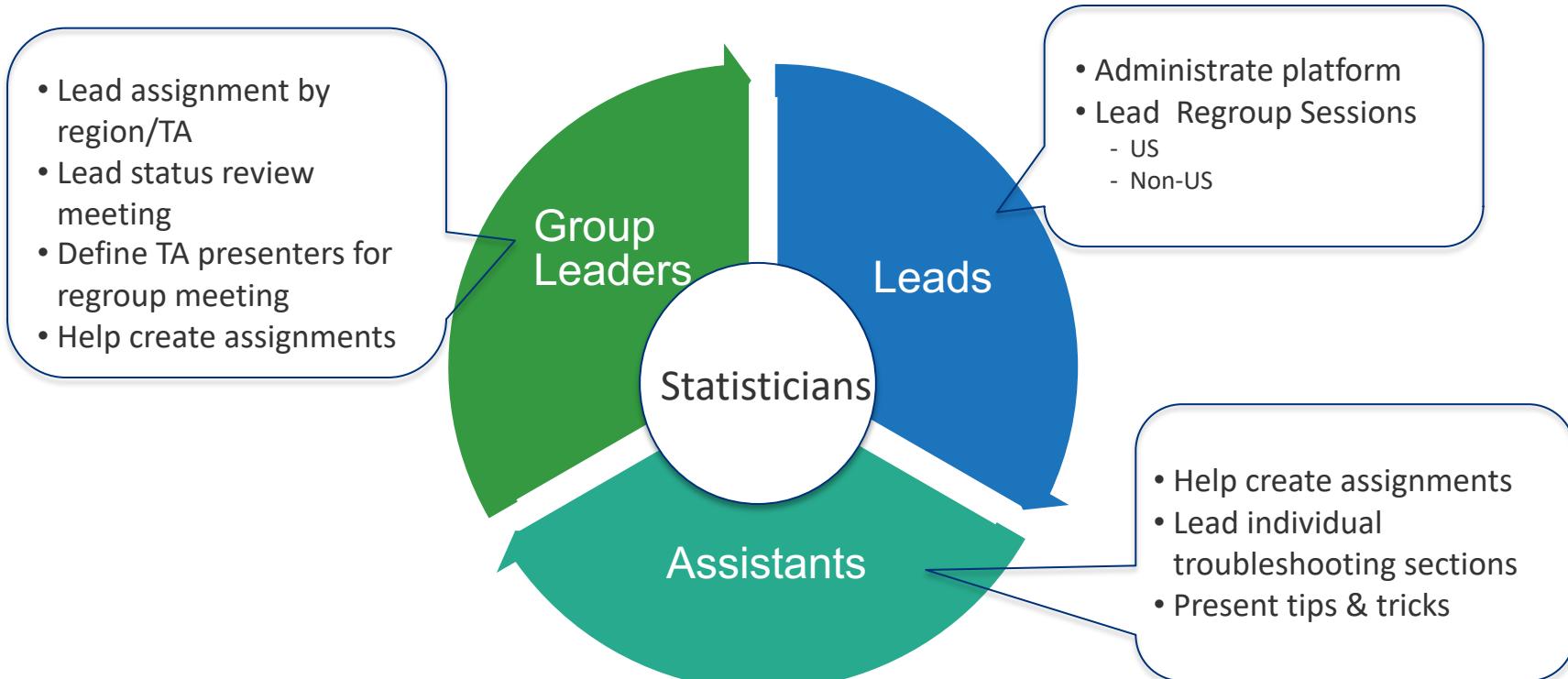
-
- 10% of what they read
 - 20% of what they hear
 - 30% of what they see
 - 50% of what they see and hear
 - 70% of what they talk over with others
 - 80% of what they use and do in real life
 - 95% of what they teach someone else
-

Crowdsource training: setting up

- Experts needed!! (~25 volunteers)
- Divide statisticians in groups (by therapeutical area/location) with 20-30 members
- Identify experts in each group volunteering to help to become group leaders
- Identify experts not assigned to a group to serve as assistants



Crowdsource training: Team roles



Assignment Workflow



RStudio.Cloud

- Why RStudio.Cloud for training?
 - Simplify the life of a learner by making it easy for them to hop on and play with R.
 - Simplify the life of a practitioner by making it easier for them to share working examples with others.
- Also show how to install R/RStudio IDE locally and how to get access to internal cloud infrastructure

RStudio.Cloud

- Create workspaces for subgroups
 - Assignment is shared with experts
- Copy project from git
- Inside project = RStudio IDE

The screenshot shows the RStudio.Cloud web interface. At the top, there's a navigation bar with 'RchitectMain' (selected), 'Projects', 'Members', and 'Info'. On the right, there are buttons for 'New Project', 'Remove', 'Options', and a search bar. Below the navigation, a sidebar on the left lists 'Spaces' (Your Workspace, RchitectMain selected), 'Learn' (Guide, What's New, Primers, DataCamp Courses, Cheat Sheets, Feedback and Questions), and 'Info' (Terms and Conditions, System Status). The main content area shows 'All Projects' with a single entry: 'First Assignment' by Stefan Avey, created on Mar 27, 2019 at 3:57 PM. It also shows a list of 'List Projects' (All, Shared with everyone, Yours) and a 'Sort Projects' section (By name, By date created). Capacity information indicates up to 23 more projects. At the bottom, there's a detailed view of the RStudio IDE environment for the 'First Assignment' project, showing the R console output, environment variables, and file browser.

RchitectMain / First Assignment Stefan Avey

R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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.
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,
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Files Plots Packages Help Viewer

Cloud project .History 0 B Mar 27, 2019, 3:58 PM
project.Rproj 205 B Mar 27, 2019, 4:44 PM
SDS_R_Graph_IntroData.xlsx 48.6 KB Mar 26, 2019, 8:26 AM
SDS_R_Graphics_IntroData_HomeWork1.txt 471 B Mar 26, 2019, 8:26 AM



Spaces

Your Workspace

RchitectMain

+ New Space

Learn

Guide

! What's New

Primer

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

All Projects

New Project



Options



First Assignment

Remove



Stefan Avey

Created Mar 27, 2019 3:57 PM



Derived from: First Assignment by Paulo Bargo

ASSIGNMENT

First Assignment



Paulo Bargo

Created Mar 26, 2019 8:25 AM

View 1 derived project ...



Delete



Move

List Projects

 All Shared with everyone Yours

Sort Projects

 By name By date created

Capacity

This space can have up to 23 more projects.

Request More Projects



Paulo Bargo



Spaces

Your Workspace

RchitectMain

+ New Space

Learn

Guide

! What's New

Primer

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
 Copyright (C) 2018 The R Foundation for Statistical Computing
 Platform: x86_64-pc-linux-gnu (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.

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 is empty

	Name	Size	Modified
<input type="checkbox"/>	..		
<input type="checkbox"/>	.Rhistory	0 B	Mar 27, 2019, 3:58 PM
<input type="checkbox"/>	project.Rproj	205 B	Mar 27, 2019, 4:44 PM
<input type="checkbox"/>	SDS_R_Graph_IntroData.xlsx	48.6 KB	Mar 26, 2019, 8:26 AM
<input type="checkbox"/>	SDS_R_Graphics_IntroData_HomeWork1.txt	471 B	Mar 26, 2019, 8:26 AM

Demo

Resources

- Central repository for all material (MS teams)
- Created a resource document: Learning Pathway document for R and Graphics
 - Online material
- RStudio.Cloud resources
- Internal videos
- Reference material

Creating data/assignments

- Start with simple graphs (using ggplot2)
 - Created a library of SAS examples (5-10) from TAs (Bar Plot, Kaplan-Meyer, Forrest Plot...)
 - Build on syllabus
- Data must be simulated/anonymized (RStudio.Cloud is public)
- Experts to rotate creating assignments
 - Data set, explanation of data, directions

First assignment

- a) XY plot progre (progesterone receptor (in fmol)) vs. estrec (estrogen receptor (in fmol)).
- b) Add Color and Shape to include information on (1) tgrade (Tumor Grade)
(2) menostat (Menopausal Status)
- c) Pairwise plots of progre (progesterone receptor (in fmol)) vs. estrec (estrogen receptor (in fmol)) vs. tsize (tumor size)
 - (it should include additional information through color and shape to provide better understanding of data)

German Breast Cancer Study Group 2

```
> data  
# A tibble: 686 x 10  
  horTh age menostat tsize tgrade pnodes progre estrec  
  <chr> <dbl> <chr>    <dbl> <chr>   <dbl> <dbl> <dbl>  
1 no     70 Post      21 II      3     48    66  
2 yes    56 Post      12 II      7     61    77  
3 yes    58 Post      35 II      9     52    271  
4 yes    59 Post      17 II      4     60    29  
5 no     73 Post      35 II      1     26    65  
6 no     32 Pre       57 III     24     0    13  
7 yes    59 Post      8 II       2    181     0  
8 no     65 Post      16 II      1    192    25  
9 no     80 Post      39 II     30     0    59
```

[Standard Solution](#)

[Group Solutions](#)

Second assignment

- Histograms, boxplot

Third assignment

- SAS dataset

⋮

Assignment 10

- RMarkdown

Learnings

- At first, it feels overwhelming
- Initially focused on graphics/visualization using ggplot2
- But, surprisingly morphed quickly
- Mandate by upper management to develop this skills



Second assignment

- Histograms, boxplot

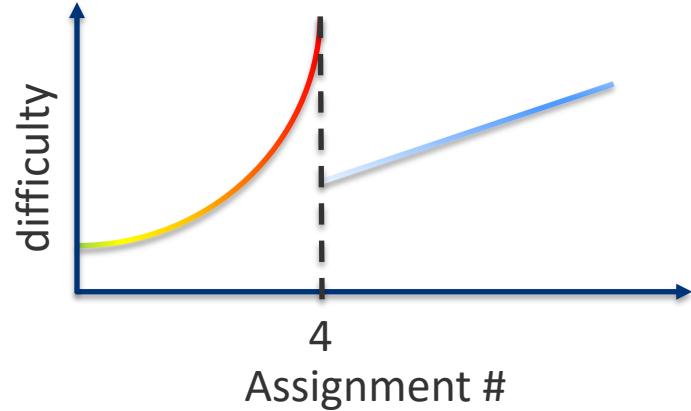
Third assignment

- SAS dataset

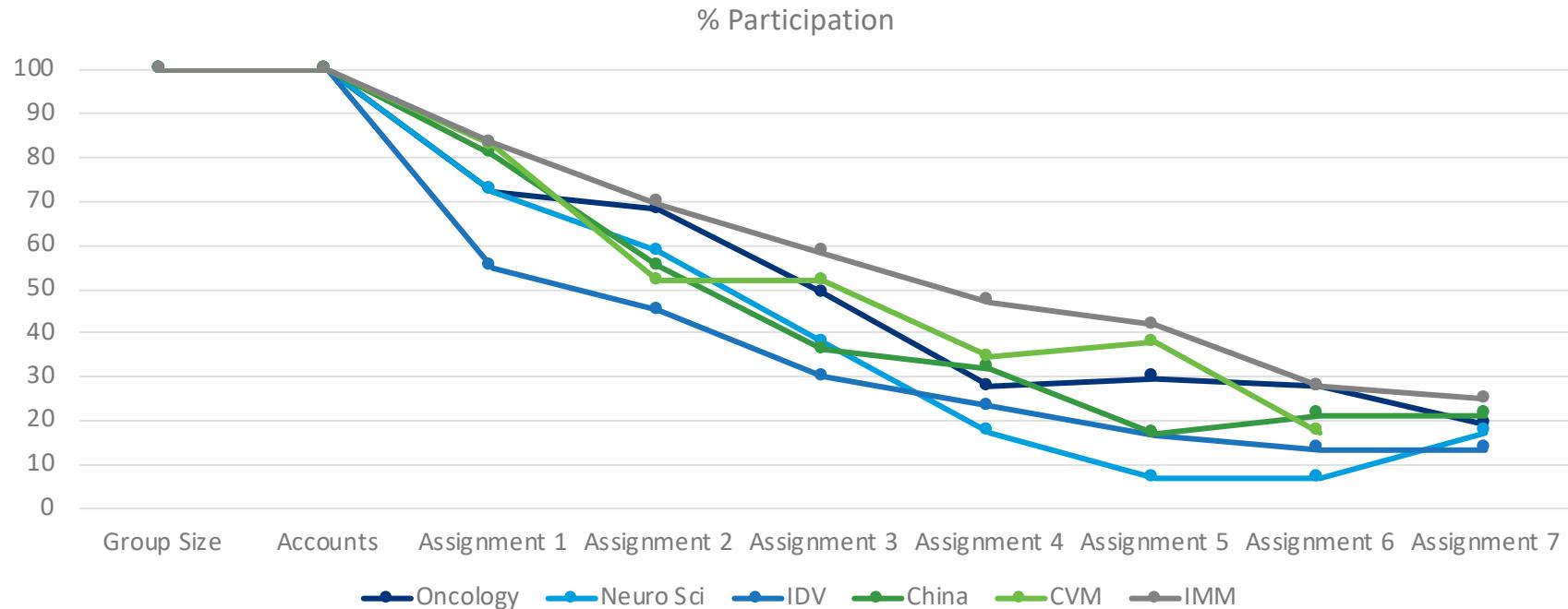
⋮

Assignment 10

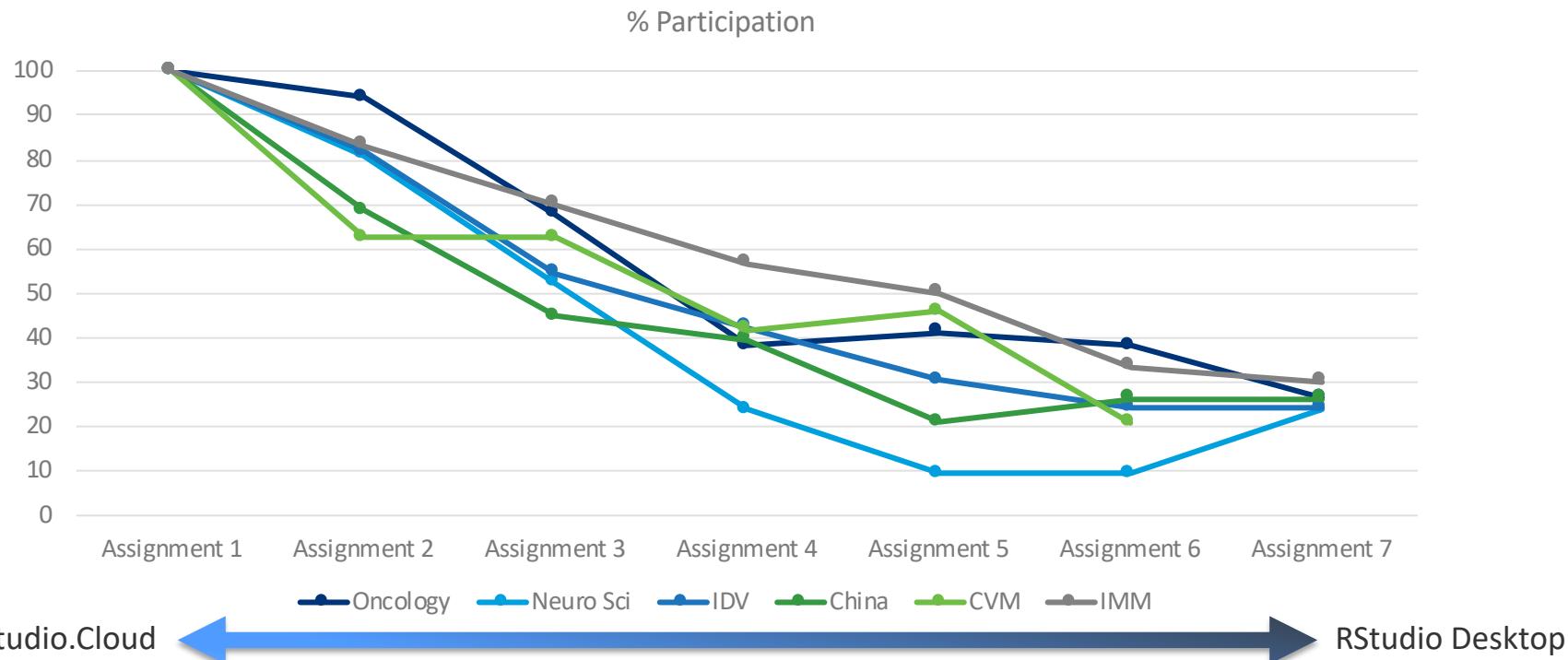
- RMarkdown



Learnings



Learnings



Conclusions

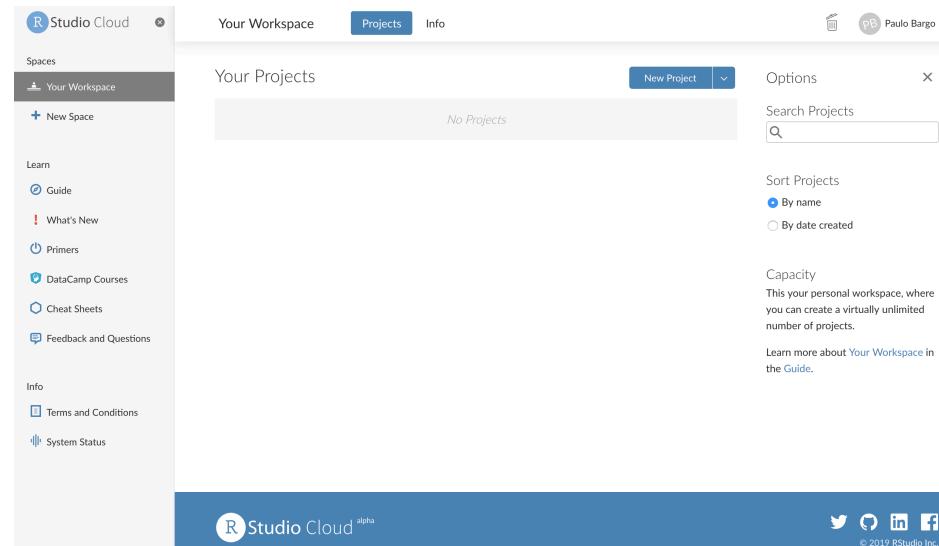
- Crowdsourcing can be an great alternative method of training
- The first steps in this journey are the hardest
- RStudio.Cloud is an excellent tool to help ease the initial hurdles
- The most import part: “The Will”

Thank you!
Questions?

Backup

Rstudio.Cloud

- Create a workspace
 - One master workspace
 - Administrated by SCO
 - All experts will be invited to join as collaborators
 - One workspace per TA
 - Administrated by TA experts
 - All TA statisticians will be invited to join as collaborators
- Workspaces are needed to share projects and control permissions



Rstudio.Cloud

- Create a project
 - One project in master workspace
 - Add instructions, data, solution, etc
 - Install packages
 - Clone and move project to TA workspace

The screenshot shows the RStudio Cloud interface. On the left is a sidebar with sections like Spaces, Learn, and Info. The main area is titled 'Rchitect Workshop / Exercise1'. It contains a 'Console' tab with a log of package installations, a 'Terminal' tab, and tabs for 'Environment', 'History', and 'Connections'. The 'Files' tab shows a list of files including 'RHistory', 'project.Rproj', 'SDS_R_Graph_IntroData.xlsx', and 'SDS_R_Graphics_IntroData_Ho...'. The 'Environment' tab shows a global environment with no objects.

```
installing binary package 'tibble' ...
* DONE [tibble]
* installing binary package 'cellranger' ...
* DONE [cellranger]
* installing binary package 'dplyr' ...
* DONE [dplyr]
* installing binary package 'forcats' ...
* DONE [forcats]
* installing binary package 'ggplot2' ...
* DONE [ggplot2]
* installing binary package 'readr' ...
* DONE [readr]
* installing binary package 'reprex' ...
* DONE [reprex]
* installing binary package 'rvest' ...
* DONE [rvest]
* installing binary package 'dbplyr' ...
* DONE [dbplyr]
* installing binary package 'haven' ...
* DONE [haven]
* installing binary package 'readxl' ...
* DONE [readxl]
* installing binary package 'tidyverse' ...
* DONE [tidyverse]
* installing binary package 'broom' ...
* DONE [broom]
* installing binary package 'modelr' ...
* DONE [modelr]
* installing binary package 'tidyverse' ...
* DONE [tidyverse]
```

The downloaded source packages are in
'/tmp/Rtmp6182lj/downloaded_packages'

Rstudio.Cloud

- Assign a project
 - Members added to TA workspace as collaborators
 - From setup wheel click on access > Make this project an assignment
 - Review projects and share screen

The screenshot shows the 'Members' tab selected in the RStudio Cloud interface. The workspace name 'Rchitect Workshop' is at the top. On the left, a sidebar lists 'Spaces' (Your Workspace, Rchitect Workshop), 'Learn' (Guide, What's New, Primmers, DataCamp Courses, Cheat Sheets, Feedback and Questions), and 'Info' (Terms and Conditions, System Status). The main area displays a table of members:

Name	Role
Bill Pikounis	Admin
Davit Sargsyan	Contributor
Paulo Bargo	Admin
Rudradev Sengupta	Admin

Below the table, there are sections for 'Options' (Invitation required, Shared), 'Access' (Add specific members to the space by sending invitations, Shared), 'Permissions' (Contributors can see the members list, Contributors can make their projects visible to all members, Viewers can see the members list), and 'Capacity' (This space can have up to 100 more members). A 'Request More Members' button is at the bottom right.