



Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

DASI: Using R Studio and Quarto for R and Python

Kellie Keeling

SEDSI, Wilmington, NC, February 16, 2023



bit.ly/KKPresentationFiles

1

Motivation

- For my Advanced Predictive Modeling with R MSBA course, I teach the course using R.
- Data Manipulation and Visuals are Nice in R
- Machine Learning is Nice in Python
- Students have requested code using Python

TENTATIVE CLASS SCHEDULE: T/Th DCB 110 6:00 - 7:50

	TOPICS		
Week 0	Install R / Complete DataCamp Courses: (Intro to Tidyverse/Intermediate R)	START HERE!>	Before Class Starts
Week 1	R/R Studio Intro/Tidyverse / Control Logic / Statistical Learning	Jan 3 (Day 1)	Jan 5 (Day 2)
Week 2	Simple/Multiple Linear Regression	Jan 10 (Day 3)	Jan 12 (Day 4)
Week 3	Regression Assumptions & Transformations / Logistic Regression	Jan 17 (Day 5)	Jan 19 (Day 6)
Week 4	Discriminant Analysis / Cross Validation/Bootstrap	Jan 24 (Day 7)	Jan 26 (Day 8)
Week 5	Ridge Regression and Lasso / Smoothers: Polynomials and Splines	Jan 31 (Day 9)	Feb 2 (Day 10)
Week 6	Gen Additive Models / CART Trees	Feb 7 (Day 11)	Feb 9 (Day 12)
Week 7	Work Week Project Part 2	Feb 14 (Zoom work hrs 6p-7:50p)	Feb 16 (No class) (No office hours)
Week 8	Bagging/Random Forest	Feb 21 (Day 13)	Feb 23 (Day 14)
Week 9	Support Vector Machines / Neural Networks	Feb 28 (Day 15)	Mar 2 (Day 16)
Week 10	Deep Learning	Mar 7 (Day 17)	Mar 9 (Day 18)
Week 11	Present Projects	Mar 14	

Sigh.... Dang Students



bit.ly/KKPresentationFiles



Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

2

R vs R Studio vs Quarto

- [R](#) From the R Project's website, "R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS."
 - R is a statistical programming language.
 - It's great for statistical programming because it was developed by statisticians
 - It's not great for general purpose programming because it was developed by statisticians.



Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

3

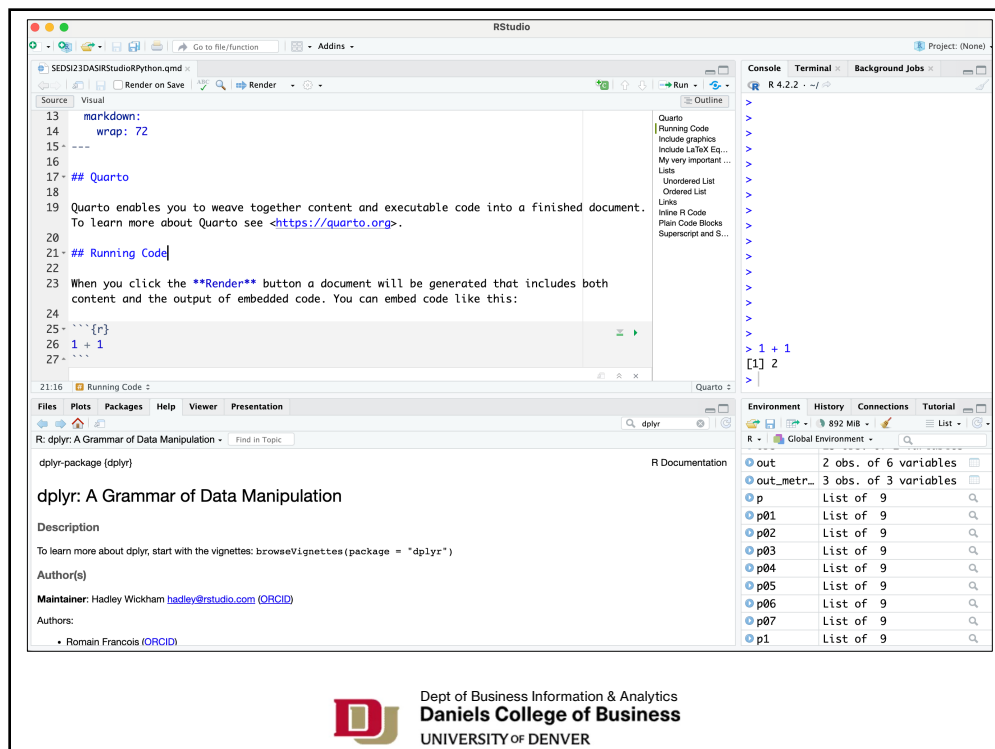
R vs R Studio vs Quarto

- [RStudio](#) From the RStudio website, "RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management." RStudio works on Mac, Windows, and linux.
 - We use the RStudio IDE in the class to write code, manage our software projects, etc. Some consider the definitive IDE to write code in R.



Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

4



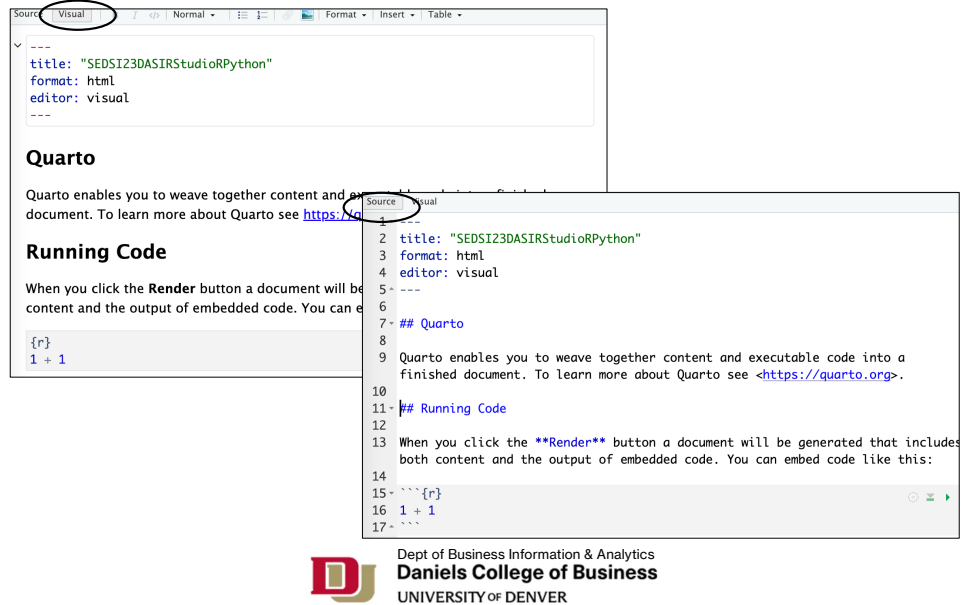
5

R vs R Studio vs Quarto

- [Quarto](#) This is an “open-source scientific and technical publishing system built on Pandoc (used for publishing workflows). Quarto can be used for R, Python, Julia, or Observable (javascript) code. This is the “latest and greatest” way to create integrated code and text documents using a “markdown” ([Markdown](#)) language (if you are familiar with Jupyter Notebooks that combine text, formatted text, code, and the output from code, this is a similar system.)

6

Visual Editor vs Source in R Studio



Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
{r}
1 + 1
```

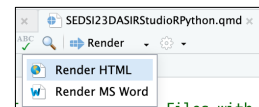
Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

7

Rendering Files

Source Code

```
16
17- ## Quarto
18
19 Quarto enables you to weave together content and executable code into a
20 finished document. To learn more about Quarto see <https://quarto.org>.
21- ## Running Code
22
23 When you click the **Render** button a document will be generated that
24 includes both content and the output of embedded code. You can embed
25 code like this:
26
27- {r}
28- 1 + 1
```



Rendered File (.html)

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

[1] 2

8

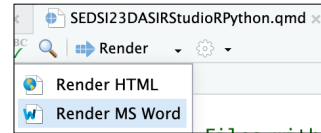
Rendering Files

Source Code

```

16
17- ## Quarto
18
19 Quarto enables you to weave together content and executable code into a
20 finished document. To learn more about Quarto see <https://quarto.org>.
21- ## Running Code
22
23 When you click the Render button a document will be generated that
24 includes both content and the output of embedded code. You can embed
25 code like this:
26
27-

```



Rendered File (.docx)

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

1 + 1

[1] 2

9

Rendered Table of Contents

Rendered File (.docx)

SEDSI23 DASI R Studio Quarto Files with R and Python		Table of contents
AUTHOR	PUBLISHED	Quarto
Kellie Keeling	February 15, 2023	Running Code
Quarto		Include graphics
Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org .		Include LaTeX Equations
		My very important paragraph
		Lists
		Links
		Inline R Code
		Plain Code Blocks
		Superscript and Subscript

Rendered File (.html)

SEDSI23 DASI R Studio Quarto Files with R and Python

Kellie Keeling
2/15/23

Table of contents

Quarto.....	1
Running Code.....	1
Include graphics.....	1
Include LaTeX Equations.....	2
My very important paragraph.....	2
Lists.....	2
Unordered List.....	2
Ordered List.....	3
Links.....	3
Inline R Code.....	3
Plain Code Blocks.....	3
Superscript and Subscript.....	3

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

← Top of Document

↑ Stays Top Right

10

Why Quarto?

(<https://quarto.org/docs/faq/rmarkdown.html>)

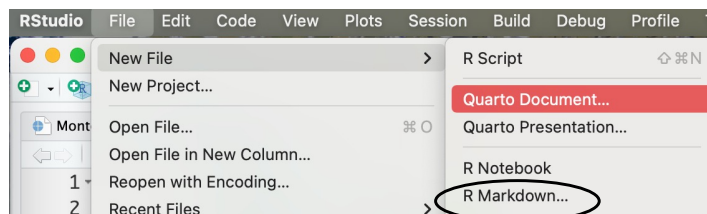
- Quarto has been gaining popularity in the last year. It has the ability to handle both R and Python in the same file.
- Quarto allows the user to keep *all* of their analysis code, output, and thoughts/comments/conclusions in a single document.
- In addition, Quarto can be used to produce reports (Word, pdf, HTML), dashboards, slides, and bookdown (create online books) among other types of output



11

Quarto Files

- How different from R Markdown?
 - .qmd extension instead of .Rmd
 - YAML (more later) at the beginning slightly different
 - It is now the top option under File, New File



- Sounds like stop enhancing .Rmd

Demoted :(



12

Markdown Basics – Headings/Lists

```
## Lists

### Unordered List

* Item 1
* Item 2
  + Item 2a
  + Item 2b

### Ordered List

1. Item 1
2. Item 2
3. Item 3
  + Item 3a
  + Item 3b

## Links
- [ROC and AUC Clearly Explained](https://www.youtube.com/watch?v=4jRBRDbJemM)
```

Line Break Important
(like before lists and before headers after code)

Lists

Unordered List

- Item 1
- Item 2
 - Item 2a
 - Item 2b

Ordered List

1. Item 1
2. Item 2
3. Item 3
 - Item 3a
 - Item 3b

Links

- [ROC and AUC Clearly Explained](https://www.youtube.com/watch?v=4jRBRDbJemM)

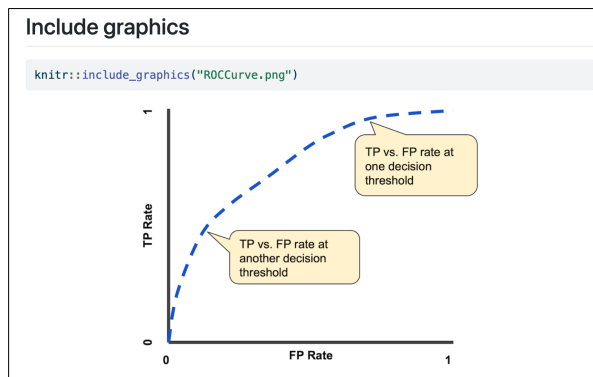


Dept of Business Information & Analytics
Daniels College of Business
UNIVERSITY OF DENVER

13

Markdown Basics – Picture Files

```
## Include graphics
```{r out.width = "60%"}
knitr::include_graphics("ROCCurve.png")
```
```



Dept of Business Information & Analytics
Daniels College of Business
UNIVERSITY OF DENVER

14

Markdown Basics – Plain, Sup/Sub

Plain Code Blocks

Plain code blocks are displayed in a fixed-width font but not evaluated

```
....
```

This is plain text

```
....
```

Superscript and Subscript

```
x^2^
```

```
x~2~
```

Plain Code Blocks

Plain code blocks are displayed in a fixed-width font but not evaluated

This is plain text

Superscript and Subscript

x^2

x_2



Dept of Business Information & Analytics
Daniels College of Business
UNIVERSITY OF DENVER

15

Quarto Basics

Quarto documents will have at least three parts.

1. **YAML** (yet another markup language) Header information that includes title, author, date, and output type (e.g., pdf or HTML) at a minimum. An example YAML section might look like the following text.

```
---
```

```
title: "SEDSI23 DASI R Studio Quarto Files with R and Python"
```

```
author: "Kellie Keeling"
```

```
date: '`r Sys.Date()`'
```

```
format:
```

```
  html:
```

```
    self-contained: true
```

```
  docx:
```

```
    default
```

```
toc: true
```

```
toc-depth: 4
```

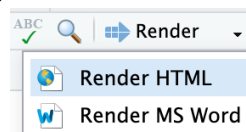
```
editor_options:
```

```
  markdown:
```

```
    wrap: 72
```

```
---
```

Important



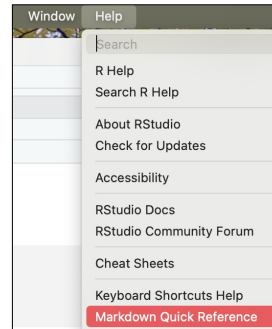
Dept of Business Information & Analytics
Daniels College of Business
UNIVERSITY OF DENVER

16

Quarto Basics

2. Markdown Code - Your section headings, lists, comments, conclusions, etc. is all written in markdown.

- There are lots of references online for learning about markdown.
- Quick Reference is under Help



Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

17

Quarto Basics

3. R Code - The actual R part of your .qmd document is embedded within what we refer to as “code chunks”.

- R can be placed inline, e.g. ``r` mean(rnorm(50))`` would evaluate the mean of 50 random normal variates within the document e.g. - 0.1685873.

```
29 Here is an example of inline R code `r` mean(rnorm(50))`
30
31 ## Include graphics
```

Here is an example of inline R code -0.1574483

Include graphics

- Or include a stand-alone chunk - if you are in the visual editor - it will just show as `{r}` but if you are in the source editor you will see three backticks as well to open and close the chunk.

Source Editor.

```
15 ` ` `{r}`
16 1 + 1
17 ` ` `
```

Visual Editor

```
{r}
1 + 1
```



Dept of Business Information & Analytics
Daniels College of Business
 UNIVERSITY OF DENVER

18

Running Python Chunks

- Find your anaconda lib location with python.exe (or python.app)

```
# Set up RStudio to run Python
```{r}
library(reticulate)
use_python("/Users/kellie.keeling/opt/anaconda3/bin")
```

# Load Python Packages
```{python}
import numpy as np
import pandas as pd
```



Dept of Business Information & Analytics  
**Daniels College of Business**  
 UNIVERSITY OF DENVER

19

## Python Read in Data

```
Read in Data
```{python}
toyotaCorolla_df = pd.read_csv('toyotacorolla.csv').iloc[:1000,:]
toyotaCorolla_df = toyotaCorolla_df.rename(columns={'Age_08_04': 'Age',
'Quarterly_Tax': 'Tax'})
toyotaCorolla_df.head()
```
```

|   | Price | Age | KM    | FuelType | HP | MetallicColor | Automatic | CC   | Doors | Weight |
|---|-------|-----|-------|----------|----|---------------|-----------|------|-------|--------|
| 0 | 13500 | 23  | 46986 | Diesel   | 90 | 1             | 0         | 2000 | 3     | 1165   |
| 1 | 13750 | 23  | 72937 | Diesel   | 90 | 1             | 0         | 2000 | 3     | 1165   |
| 2 | 13950 | 24  | 41711 | Diesel   | 90 | 1             | 0         | 2000 | 3     | 1165   |
| 3 | 14950 | 26  | 48000 | Diesel   | 90 | 0             | 0         | 2000 | 3     | 1165   |
| 4 | 13750 | 30  | 38500 | Diesel   | 90 | 0             | 0         | 2000 | 3     | 1170   |



Dept of Business Information & Analytics  
**Daniels College of Business**  
 UNIVERSITY OF DENVER

20

## Python Regression

```
```{python}
# create a fitted model
sm_reg = smf.ols(formula='Price ~ Age + Doors + Automatic',
data=toyotaCorolla_df)
sm_reg_results = sm_reg.fit()
sm_reg_results.summary()
```
```

| OLS Regression Results |                  |         |                     |           |          |          |
|------------------------|------------------|---------|---------------------|-----------|----------|----------|
| Dep. Variable:         | Price            |         | R-squared:          | 0.759     |          |          |
| Model:                 | OLS              |         | Adj. R-squared:     | 0.758     |          |          |
| Method:                | Least Squares    |         | F-statistic:        | 1044.     |          |          |
| Date:                  | Sat, 23 Apr 2022 |         | Prob (F-statistic): | 6.61e-307 |          |          |
| Time:                  | 16:29:17         |         | Log-Likelihood:     | -8936.7   |          |          |
| No. Observations:      | 1000             |         | AIC:                | 1.788e+04 |          |          |
| Df Residuals:          | 996              |         | BIC:                | 1.790e+04 |          |          |
| Df Model:              | 3                |         |                     |           |          |          |
| Covariance Type:       | nonrobust        |         |                     |           |          |          |
|                        | coef             | std err | t                   | P> t      | [0.025   | 0.975]   |
| Intercept              | 2.033e+04        | 333.558 | 60.958              | 0.000     | 1.97e+04 | 2.1e+04  |
| Age                    | -193.1023        | 3.545   | -54.478             | 0.000     | -200.058 | -186.146 |
| Doors                  | 185.4537         | 62.072  | 2.988               | 0.003     | 63.646   | 307.261  |
| Automatic              | 961.6982         | 272.807 | 3.525               | 0.000     | 426.356  | 1497.040 |

21

## Extracting R code from Quarto

- You can extract all of the R code from a quarto markdown file and put it into an .R file using the `purl()` function from the `knitr` package.
  - You can run from console.
  - You can run within a chunk.
- Recommend liberal use of comments since all text will be lost.

```
Extracting R Code
```{r}
knitr::purl(input="SEDSI23DASIRStudioRPython.qmd",
output = "YOURPATH/SEDSI23DASIRStudioRPython.R")
```
```

```
SEDSI23DASIRStudioRPython.R

1 + 1

-----out.width = "60%"-----
knitr::include_graphics("R0CCurve.png")

knitr::purl(input="SEDSI23DASIRStudioRPython.qmd", output = "YOUR PATH/SEDSI23DASIRStudioRPython.R")
```

22



**KKeeling@du.edu**



[bit.ly/KKPresentationFiles](https://bit.ly/KKPresentationFiles)



Dept of Business Information & Analytics  
**Daniels College of Business**  
UNIVERSITY OF DENVER

23