



**COLORADO STATE
UNIVERSITY**

Fast Facts: Reproducible Fact Book Reporting in Quarto (RStudio)

AIR Forum 2025
Wednesday May 21st 2:45 – 3:30PM

Nicole Ross, Associate Director
nmvross@colostate.edu

Lee Tyson, Senior Research Analyst
lkytson@colostate.edu

Presentation Objectives

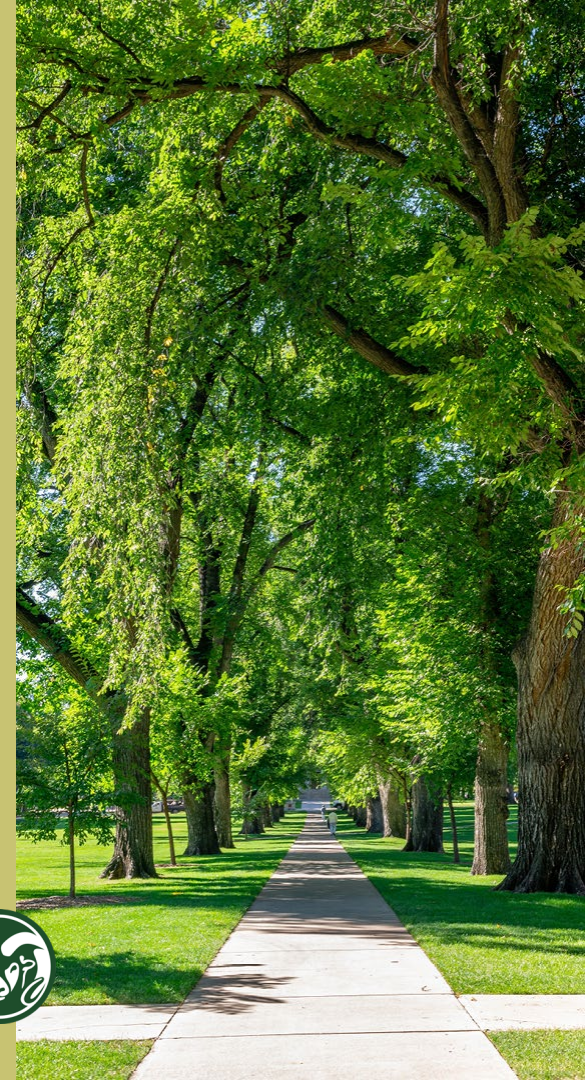


- Explore how Quarto (using RStudio) can function as an open-source **complement** and **substitute** to other reporting tools for **IR offices of all sizes, budgets, and reporting demands.**
- Show the coding mechanics in RStudio that transform source files into professional deliverables via Quarto through **live demonstrations**
- Introduce the RStudio and Quarto interfaces and provide a **Quarto Starter Kit** guide
- **Assuming:** novice/beginner knowledge of RStudio/Quarto, but some familiarity with or exposure to programming or coding (VBA, DAX, R, Python, etc.)



About Colorado State University

- Type: **Public land-grant university**
- Classification: **R1**
- Founded: **1870**
- Location: **Fort Collins, Colorado** (Pop. 170,000)
- Enrollment: **32,000+ students** (UG 25,000, GR 7,200)
- Employees: **11,000 faculty and staff members**
- Academic Program: **279 degrees**
(74 bachelors, 140 masters, 65 doctoral)



IR Context

- **Institutional Research, Planning and Effectiveness**
 - 7 IR staff members
 - IR-specific, Oracle on-prem database
 - 35 Power BI interactive dashboards
 - Familiar with programming languages
 - PL/SQL
 - Stata, SPSS, R, Python
 - DAX, Power Query



PL/SQL
Commands



Limitations of IR Reporting & Analysis Tools



Business Intelligence Tools

- **Cost:** Enterprise licenses can be prohibitively expensive
- **Complexity:** Require significant campus-wide resources and training
- **Confinement:** Limited report accessibility to external stakeholders (data security)

Analysis Tools

- **Brittle:** Can break or malfunction with minor or unexpected changes to data, formulas, etc.
- **Bespoke:** Not easy to reproduce or automate output for other use cases

Statistical Programming Tools

- **Atomized:** The analysis and the narrative are separate, fragmented and need to be compiled

Hello, Quarto! What's RStudio? Posit?



An open-source scientific and technical publishing system

- Creates dynamic documents, websites, presentations
- **Free** and accessible
- Integrates with R, Python, Julia, Observable
- **Reproducible** and **automated** workflows
- **Multiple output formats** (PDF, Word, Excel, HTML, etc.)
- **Underlying data** are not exposed



Company



IDE
(Integrated
Development
Environment)



Publishing
System

Back End – to – Front End



```
---  
title: "ggplot2 demo"  
author: "Norah Jones"  
date: "5/22/2021"  
format:  
  html:  
    code-fold: true  
---  
  
## Air Quality  
  
@fig-airquality further explores the impact of temperature on ozone level.  
  
``{r}  
#| label: fig-airquality  
#| fig-cap: "Temperature and ozone level."  
#| warning: false  
  
library(ggplot2)  
  
ggplot(airquality, aes(Temp, Ozone)) +  
  geom_point() +  
  geom_smooth(method = "loess")  
---
```

ggplot2 demo

Norah Jones
May 22nd, 2021

Air Quality

[Figure 1](#) further explores the impact of temperature on ozone level.

► Code

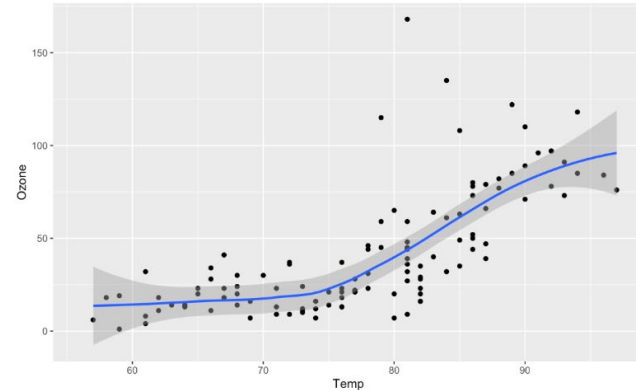


Figure 1: Temperature and ozone level.

Maslow's Hammer



The over-reliance on a particular tool simply because that tool is either more immediately available or because it's more familiar.

“IF THE ONLY TOOL YOU HAVE IS A HAMMER,
IT IS TEMPTING TO TREAT EVERYTHING
AS IF IT WERE A NAIL”



Abraham Maslow
The Psychology of Science, published in 1966

Maslow's Hammer

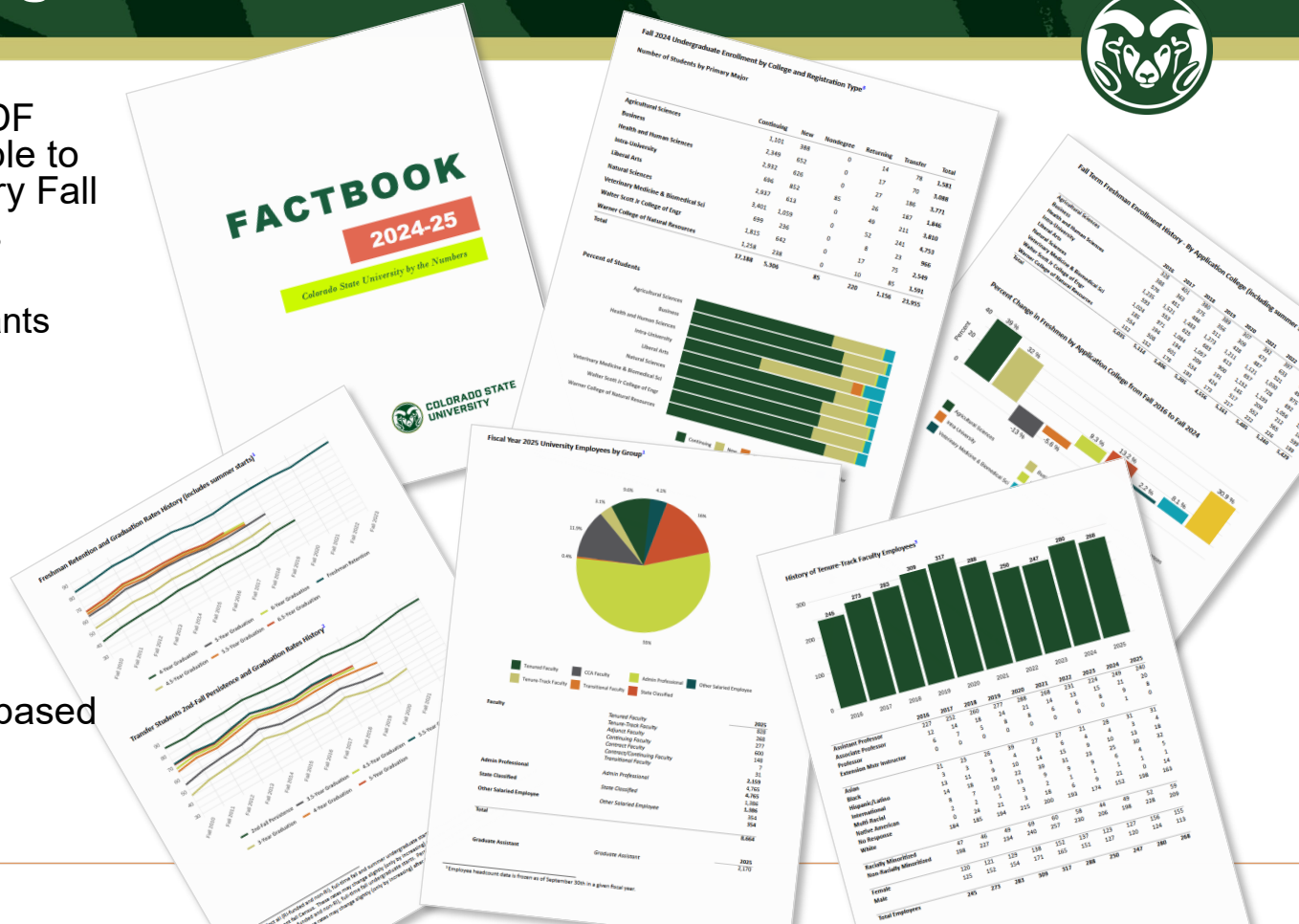


- Quarto is an excellent **complement** and **substitute** reporting tool for IR offices of all sizes, budgets, and reporting demands.
- Especially excellent for:
 - **Static** (as opposed to **interactive**) reports or summaries
 - Reports, presentations, datasets, or **processes** that need to be replicated on a **regular basis** (weekly, annual, etc.)
 - Deliverables that need to be **reproduced**, but with **parameters** changed slightly (the college or department, the academic year, etc.)
 - Reports meant for **external stakeholders**

It began with our Fact Book...



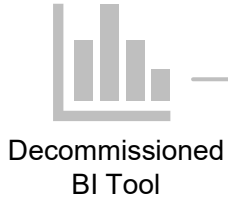
- **Fact Book** – a static PDF document made available to external audiences every Fall
 - Board of Governors
 - Legislators
 - Prospective Applicants
- Displays key facts on
 - Enrollment
 - Entering Freshman
 - Credit Hours
 - Student Success
 - Employees
- Transitioned to Quarto-based version in 2022



Fact Book Creation Process



Old Approach



Navigate to BI report of interest (e.g., Enrollment)

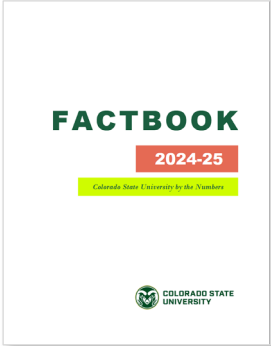
Select filters and sorts to display intended info (e.g., Fall 2024 Residents)

Brittle
Bespoke
Atomized
Time-intensive
Old BI tool

Alternative 1



PL/SQL Commands



Alternative 2



Robust
Reproducible
Integrated
Automated
Dynamic
Documented
STILL initially time-intensive
Learning curve

Fact Book Creation in Quarto



Alternative 2

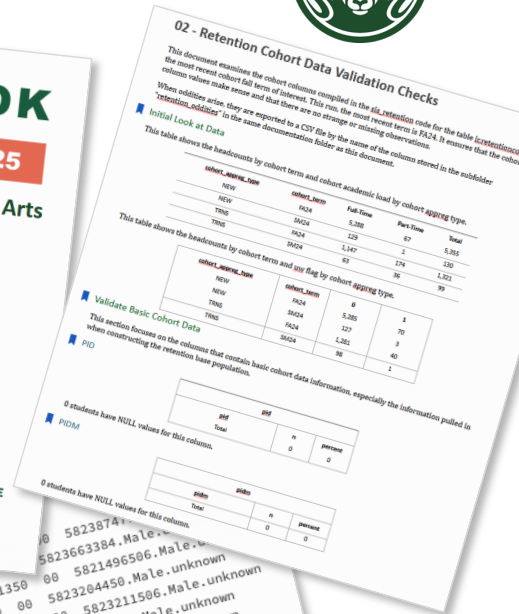


- Contracted and collaborated with Geospatial Centroid at CSU
- 9-month process
- Frequent meetings and knowledge transfer sessions
 - Section-by-section chunks
 - Ran new process in parallel with “old”
- Mutually beneficial (Quarto and RStudio for us, but also PL/SQL for them)
- **CSU IR leadership strongly encouraged us to take this approach, valuing the skill development and future efficiencies to be gained**

Quarto in Action



- Fact Book
- College-Level Fact Books
- Data Validation Reports
- National Student Clearinghouse



IR_30_GR_m_unknown.txt	Colorado State University	20241222
1 HI 001350 0	20000322	20190615
2 D1	20000613	20190615
3 D1	20000513	20190615
4 D1	20000112	20190615
5 D1	20000711	20190615
6 D1	20000110	20190615
7 D1	19990830	20190615
8 D1	19930928	20190615
9 D1	20000706	20190615
10 D1	20000903	20190615
11 D1	19991104	20190615
12 D1		
TT 13		



General Layout Introduction

WHAT IS QUARTO/RSTUDIO



Example 1 – Enrollment Report

FACT BOOK



Important for efficiency and reproducibility

FOLDER STRUCTURE



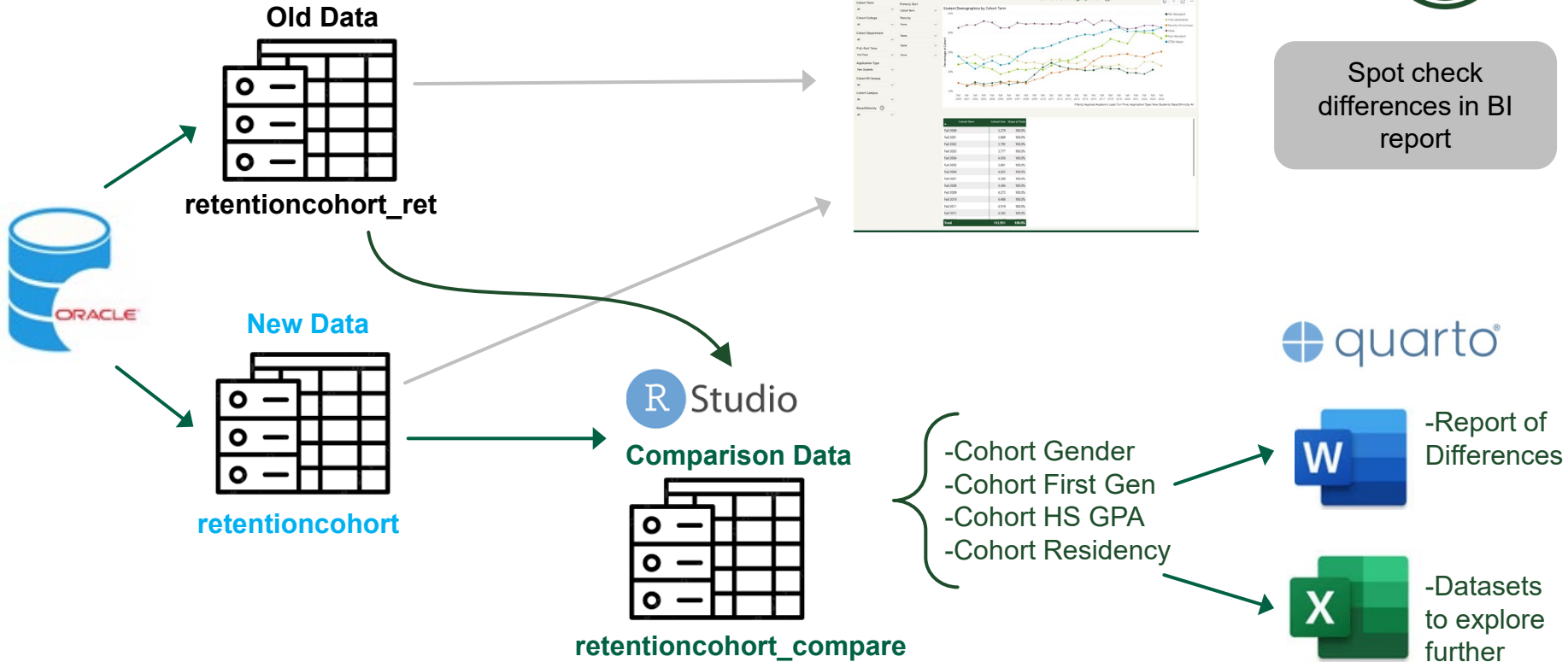
Example 2 - Parameters

COLLEGE-LEVEL FACTBOOKS



RETENTION DATA VALIDATION

Retention Data Validation



Retention Data Validation



Cohort residency category desc

```
{r}
compare_tab(retentioncohort_compare
, 'cohort_residency_category_desc_new', 'cohort_residency_category_desc_old'
, filter_to_na = FALSE)

diff_residency_desc <- diff_count(retentioncohort_compare,
'cohort_residency_category_desc_new',
export = TRUE)
```

`r diff_residency_desc$diff_ref` students have different values between the new retention and old sis_retention.

Here is a cross-tab between the overhauled and old sis_retention.

```
{r}
crosstab(retentioncohort_compare, 'cohort_residency_category_desc_new',
'cohort_residency_category_desc_old')
```

Cohort residency category desc

cohort_residency_category_desc	n	percent_new	n_old	percent_old	diff
Non-Resident	59,993	28.3	60,016	28.3	-23
Resident	151,787	71.7	151,764	71.7	23

29 students have different values between the new retention and old sis_retention.

Here is a cross-tab between the new and old sis_retention.

cohort_residency_category_desc_old		
cohort_residency_category_desc_new	Non-Resident	Resident
Non-Resident	59,990	3
Resident	26	151,761
Total	60,016	151,764

13	SM13	FA13	TRNS	Full-Time	M	0	Resident	Non-Resident		
14	SM12	FA12	TRNS	Part-Time	M	0	Resident	Non-Resident		
15	SM11	FA11	TRNS	Full-Time	M	0	Resident	Non-Resident		
16	SM11	FA11	NEW	Full-Time	M	0	Non-Resident	Resident		
17	SM10	FA10	TRNS	Full-Time	M	0	Resident	Non-Resident		
18	SM06	FA06	TRNS	Full-Time	M	0	Resident	Non-Resident		
19	SM02	FA02	TRNS	Full-Time	M	0	Resident	Non-Resident		
20	SM02	FA02	TRNS	Part-Time	M	1	Resident	Non-Resident		
21	SM02	FA02	TRNS	Full-Time	M	0	Resident	Non-Resident		
22	SM01	FA01	TRNS	Full-Time	M	0	Resident	Non-Resident		
23	SM99	FA99	TRNS	Full-Time	M	0	Resident	Non-Resident		
24	SM99	FA99	TRNS	Full-Time	M	0	Resident	Non-Resident		
25	SM98	FA98	TRNS	Full-Time	M	0	Resident	Non-Resident		
26	SM97	FA97	TRNS	Full-Time	M	0	Resident	Non-Resident		
27	SM95	FA95	TRNS	Full-Time	M	0	Resident	Non-Resident		
28	SM95	FA95	TRNS	Full-Time	M	0	Resident	Non-Resident		
29	SM93	FA93	TRNS	Part-Time	M	0	Resident	Non-Resident		



NATIONAL STUDENT CLEARINGHOUSE

Prepping Data for NSC Cohort Report for IPEDS GR



Data prepped in SQL
includes codes for
subgroups (e.g., male
Latinx, female Asian)
(30-31)

Recode groups
with n < 11
(41-55)

```
30 # pull in IR data
31 nsc_cohort_pre <- dbGetQuery(con, statement = read_file('GR sql pull.sql'))
32
33 # get the counts of the gender/race combinations
34 # and recode as needed in next step
35
36 ( group_count <- nsc_cohort_pre |>
37   group_by( file_group, file_group_desc ) |>
38   summarize( cnt = n() )
39 )
40
41 # add labels to the filter group and combine file groups with N < 11
42 # these recods will be the basis for the exported text files to be
43 # uploaded to NSC
44
45 nsc_cohort <- nsc_cohort_pre |>
46   mutate(
47     file_group2 = case_when(
48       file_group == 4 | file_group == 7 | file_group == 16 | file_group == 18 ~ 99,
49       .default = file_group
50     ),
51     file_group2_desc = case_when(
52       file_group == 4 | file_group == 7 | file_group == 16 | file_group == 18 ~ 'combo',
53       .default = file_group_desc
54     )
55   )
56
57 # updated headcounts by subgroups, these numbers are used/modified in the
58 # export as the T1 (terminator) line of the text files
59
60 ( group_count2 <- nsc_cohort |>
61   group_by( file_group2_desc ) |>
62   summarize( cnt = n() )
63 )
```

Calculate headcounts in
each subgroup
(33-39)

Check headcounts
on recoded groups
(57-63)

Generate files for Upload to NSC



Generate list values
named `group_list` (65)

Function named
`ncs_files` that generates
a tab-delimited text file;
includes header,
student data, and trailer
row
(72-94)

```
65 group_list <- as.list( group_count2$file_group2_desc)
66
67 # this is the header that begins each text file for upload
68 # update date; note "CO" stands for cohort report, not Colorado!
69 header <- data.frame( "H1", "001350", "0", "Colorado State University", "20241222", "CO", "I")
70
71
72 # define function that generates the file for the specific subgroup
73 nsc_files <- function( sub_group ){
74
75   header |>
76     write.table( file = paste0('upload/IR_30_GR_', sub_group, '.txt'),
77                 quote = FALSE, row.names = FALSE, col.names = FALSE, sep = "\t" )
78
79   nsc_cohort |>
80     filter( file_group2_desc == paste0( sub_group ) ) |>
81     select ( -contains( 'file' ) ) |> # drops the file_group variables
82     write.table( file = paste0('upload/IR_30_GR_', sub_group, '.txt'),
83                 quote = FALSE, row.names = FALSE, col.names = FALSE, sep = "\t", append = TRUE, na = '' )
84
85   group_count2 |>
86     filter( file_group2_desc == paste0( sub_group ) ) |>
87     mutate(
88       file_group2_desc = 'T1',
89       cnt = cnt + 2 # to account for the header and terminator lines
90     ) |>
91     write.table( file = paste0('upload/IR_30_GR_', sub_group, '.txt'),
92                 quote = FALSE, row.names = FALSE, col.names = FALSE, sep = "\t", append = TRUE )
93
94   }
95
96 # call the function that will generate the files to be uploaded by looping through all the values
97 # in the group_list
98
99 for ( x in group_list ) {
100
101   group_list[[x]] <- nsc_files( x )
102
103 }
```

Create header per
NSC specifications
(67-69)

Function that loops
through `nsc_files` for
each value in
`group_list` and writes
text file for each one
(96-103)

Keep in Mind



Getting Started

- **We have a Guide!**
- Download R, RStudio, & Quarto*
- Folder structure matters
- Start small

Tech Considerations

- On-prem vs. cloud storage for shared files and data
- Consider version control

Programming

- Writing code is essential, but resources exist
- Tidyverse syntax
- Packages

Remember...

- You don't need to be an expert
- Some opacity is okay
- Consult resources, including AI

Consider Using

- **Static** deliverables
- **Replicated** reports, processes, tasks, etc. on **regular** basis (weekly, annually, etc.)
- The same output, but with **parameters** slightly tweaked (college, year, etc.)
- **External stakeholder** requests

Resources From Today on Eventscribe



- **Presentation Slides**
- **Example 1 – Fall Enrollment**
 - The PDF output
- **Quarto Starter Kit**
 - How do download R, RStudio, Quarto
 - How to get started
 - More Examples
 - Best Practices
- To download the source .qmd files for Example 1 – Fall Enrollment and the Quarto Starter Kit, please go here:
<https://github.com/nmvross/airforum2025/tree/main>



airforum2025 Public Pin Unwatch 1

main 1 Branch 0 Tags Add file <> Code

nmvross Add Quarto Starter Kit.md ecfac7e · now 5 Commits

LICENSE	Initial commit	31 minutes ago
README.md	Add Quarto Starter Kit.md	now
ex1-fall-enrollment.qmd	Uploading .pdf and .qmd to share	28 minutes ago
ex1-fall-enrollment.pdf	Uploading .pdf and .qmd to share	28 minutes ago
quarto-starter-kit.pdf	Upload quarto starter kit resources	4 minutes ago
quarto-starter-kit.qmd	Upload quarto starter kit resources	4 minutes ago

README MIT license

airforum2025

Resources for AIR Forum 2025

The ex1-fall-enrollment.qmd file can be downloaded and opened in RStudio/Quarto to render the accompanying ex1-fall-enrollment.pdf file. Play around with the .qmd code to render your own .pdfs!

The quarto-starter-kit.qmd file can also be downloaded and opened in RStudio/Quarto to render the accompanying .pdf file. This may have additional tips and tricks to see within RStudio/Quarto, and serves as an additional example of how you can produce a deliverable from within RStudio.



**COLORADO STATE
UNIVERSITY**

Q&A

Nicole Ross | nmvross@colostate.edu

Lee Tyson | lktyson@colostate.edu



Empowering Student Success Through Technology-Driven Interactive Dashboards

Learn how Colorado State University has leveraged interactive dashboards to support stakeholders in making data-informed decisions

Tae Yamaki, Heather Novak & Nicole Ross

Thursday, May 22

9:45-10:30 am

Sebastian L1+2

