



Institute  
and Faculty  
of Actuaries

OCTO

# Putting data science to work – a case study

Alex Breeze & Martin Tynan  
Octo Telematics

5 November 2018

# Putting data science to work

## Setting the scene



5 November 2018

3

## Version control using git



5 November 2018

8

## Managing dependencies in R and Python



5 November 2018

18

## Final thoughts



5 November 2018

95

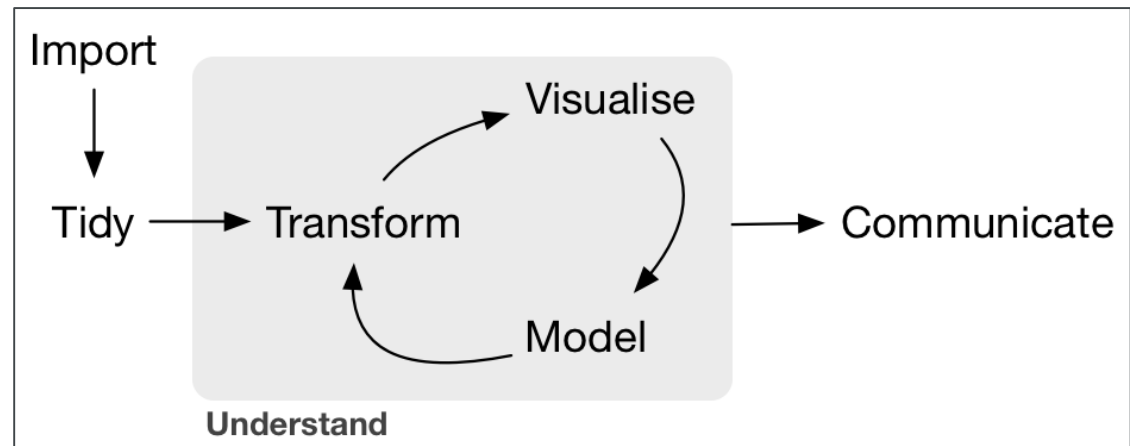
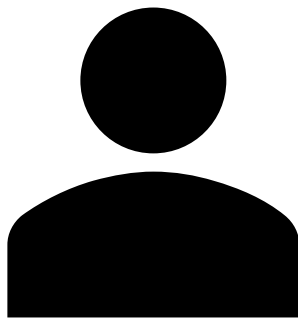


Institute  
and Faculty  
of Actuaries

# Setting the scene



Institute  
and Faculty  
of Actuaries

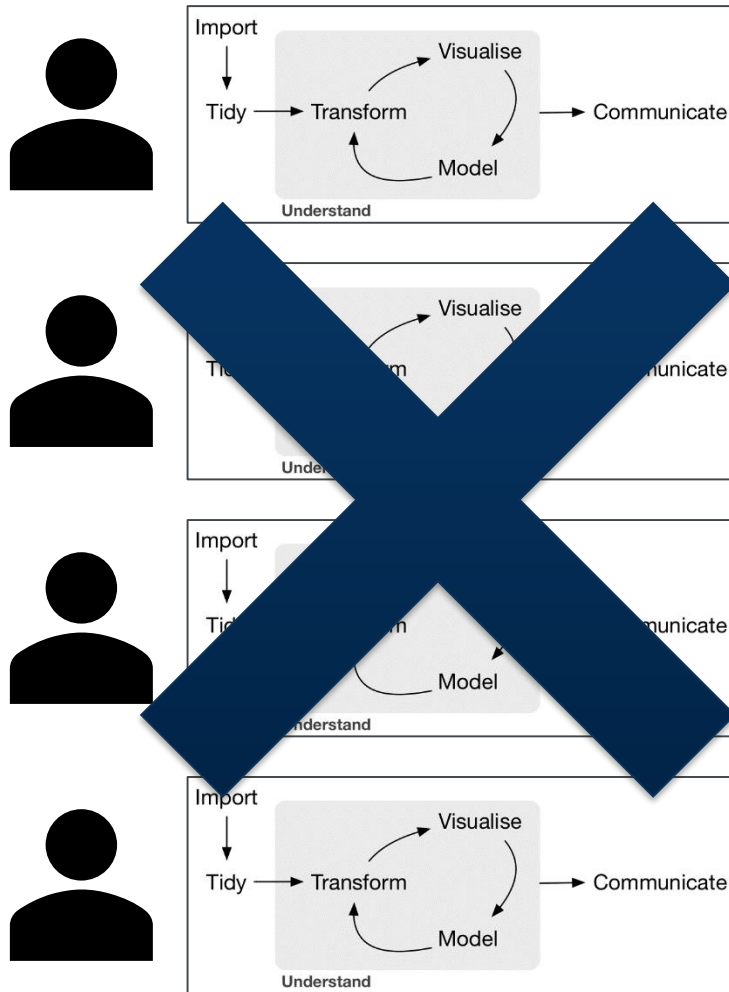


Data science workflow picture from: <https://ismayc.github.io/poRtland-bootcamp17/>



Institute  
and Faculty  
of Actuaries

# Data science team?



## Differences

Collaboration	Review
Documentation	Reproducibility



## Results

- Reliability of solution
- Sustainably** adding value
- Regulatory approval



Institute  
and Faculty  
of Actuaries

# What's new? Open source!

Opportunities	Risks
Cutting edge Rapid improvements	Is it trustworthy? Rapidly outdated
Free / open licence Leverage existing solution	Use at own risk Learning curve
New tools to <i>embed</i> good practice	New concepts



# This presentation

1. Version control using Git
2. Managing dependencies in R and Python



Our experiences

## Caveat:

- This is not (and will never be) the final version of these slides!
- Links to resources at end.

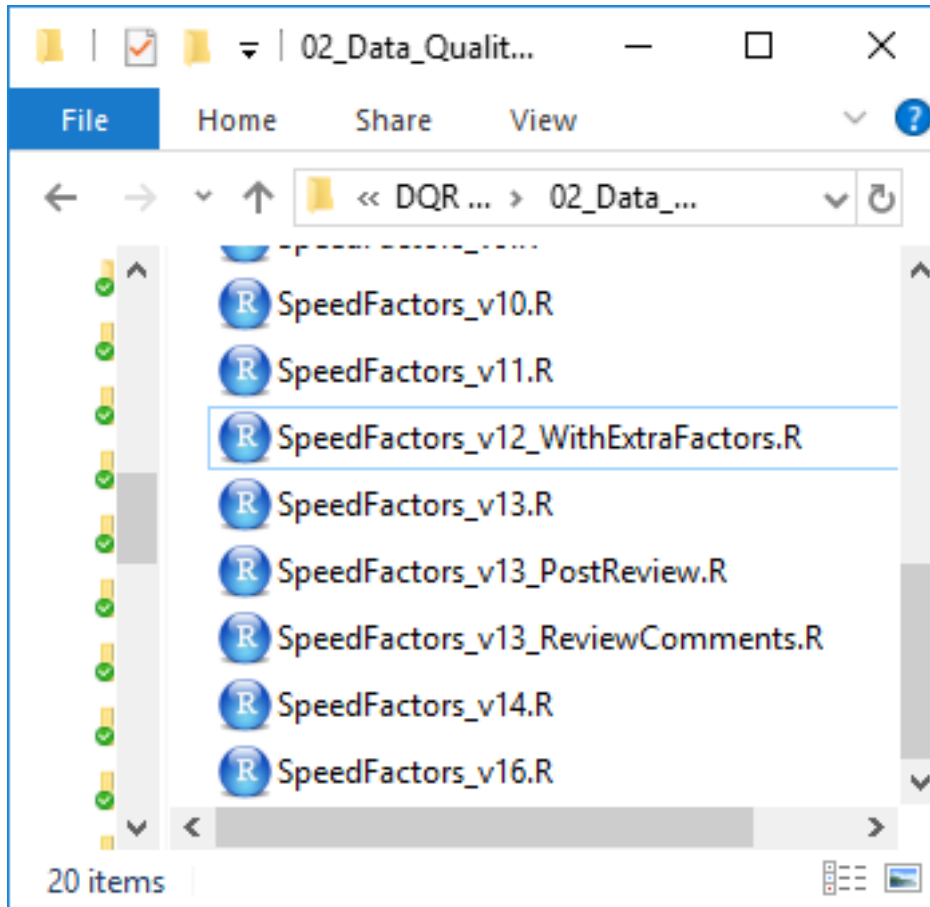


Institute  
and Faculty  
of Actuaries





# Version Chaos



## Why?

- Experimenting
- Feedback
- Reproducibility

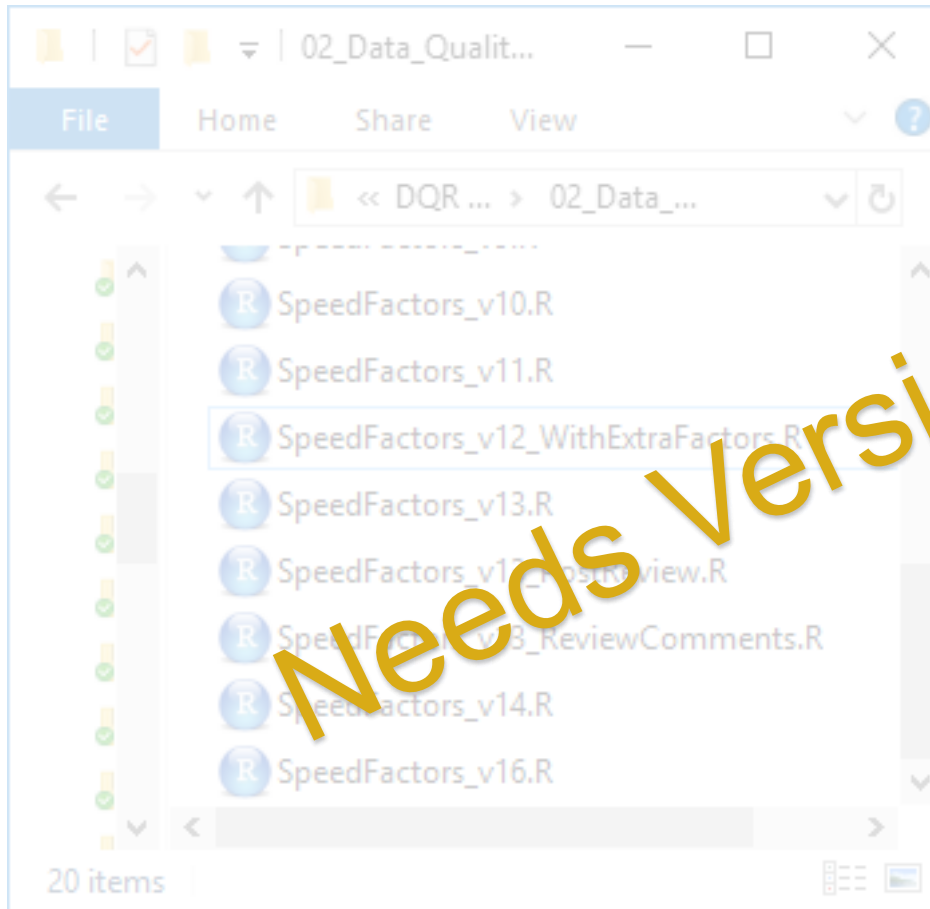
## How do you choose?

- File name
- Date modified
- Look at each file
- Documentation



Institute  
and Faculty  
of Actuaries

# Version Chaos



## Why?

- Experimenting
- Feedback
- Reproducibility

## How do you choose?

- File name
- Date modified
- Look at each file
- Documentation



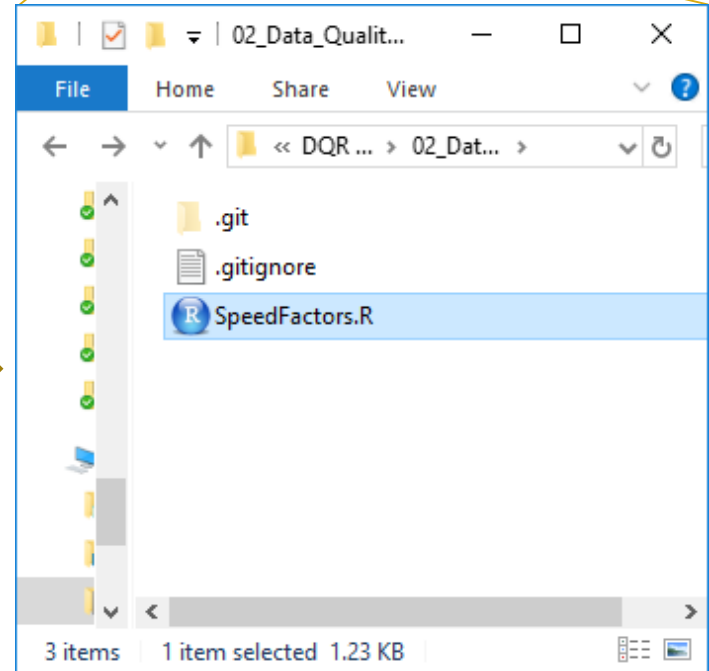
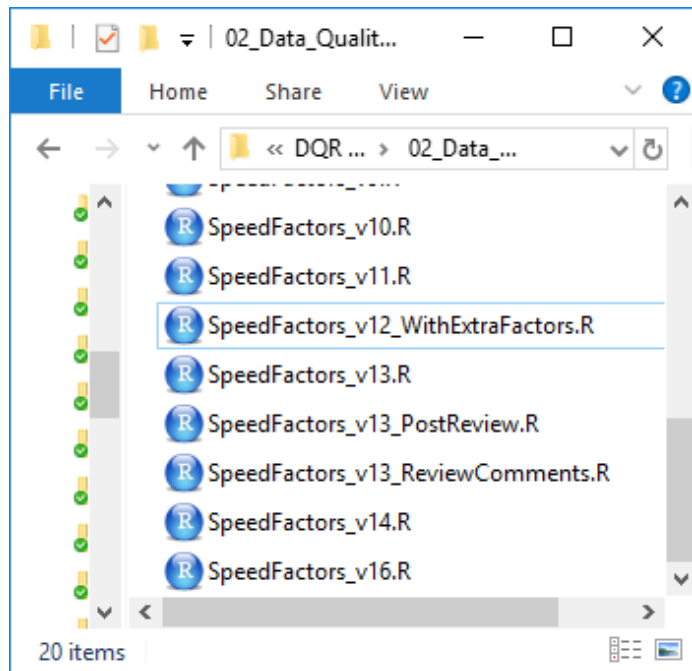
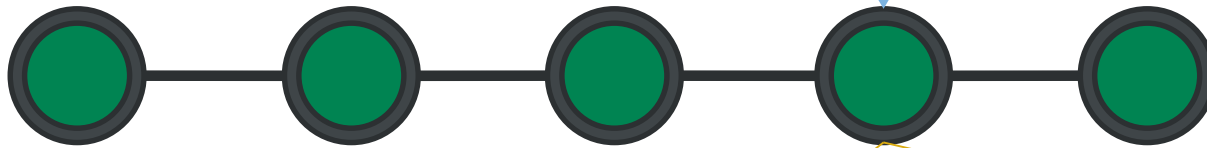
Institute  
and Faculty  
of Actuaries

# Git

Snapshot of project & comment

*Tree of commits*

Project beginning

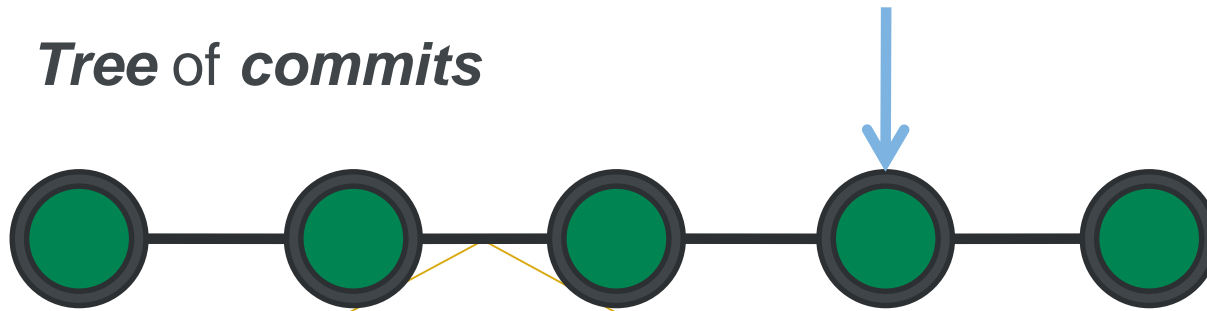


# Git

Snapshot of project & comment

*Tree of commits*

Project beginning



```
SHA      eb197822
Author   Martin Tynan <martin.tynan@octotelematics.com>
Date     2018-08-23 09:03
Subject   Automatically installs phantomJS if it can't be found.
Parent   101e9951

200 Code/global.r

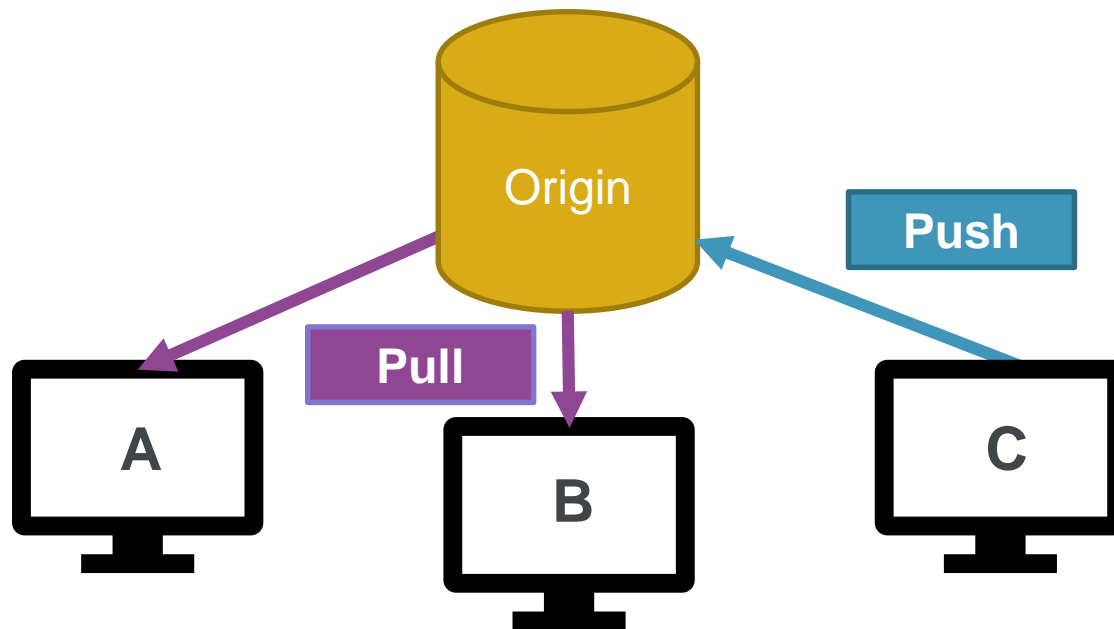
200 Code/global.r
@@ -14,6 +14,9 @@ library(tidyr)
14 14 source.DQR <- "../DriveAbility Functions 2016-10-24.r"
15 15 source(source.DQR, echo = FALSE)
16 16
17 17 if(is.null(webshot::find_phantom())) {
18 18   webshot::install_phantomjs()
19 19 }
17 20
18 21 # Icons
19 22 journeyMapLocationIcon <- makeIcon(iconUrl =
```



Institute  
and Faculty  
of Actuaries

# Git in a Team

Need to *share* the *complete* project code folders

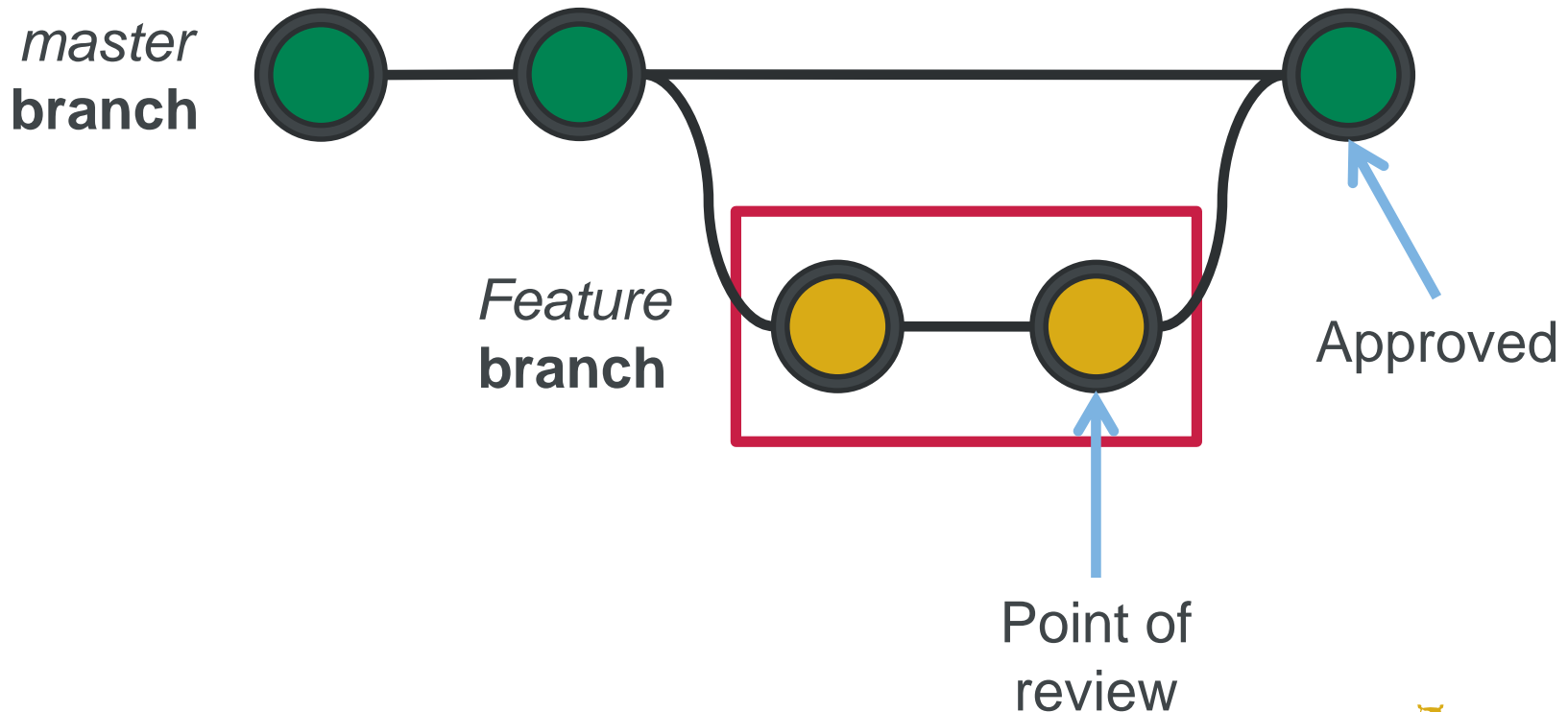


You are in control of when the syncing occurs

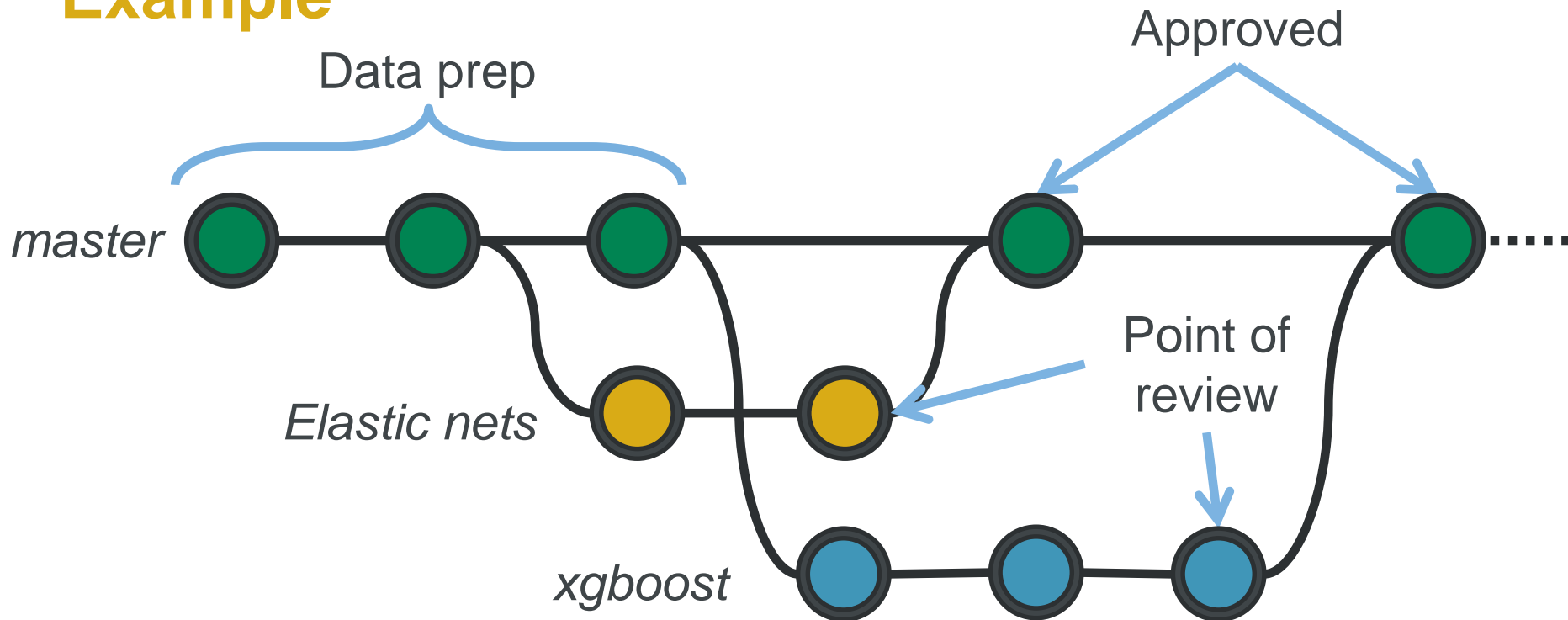


Institute  
and Faculty  
of Actuaries

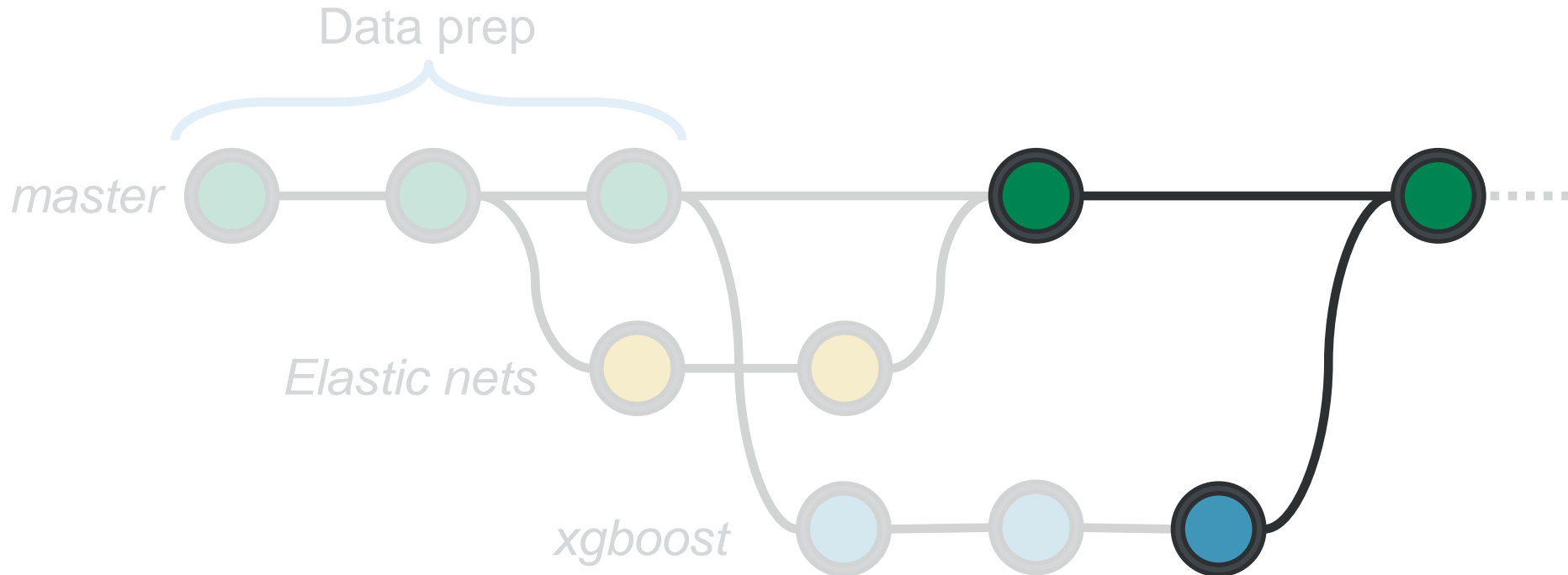
# 'Feature Branch' Workflow



# Example



# Example



**Merging** is where you combine two commit trees into one unified history.

This can lead to conflicts!



Institute  
and Faculty  
of Actuaries



# Considerations

- **Tracks** projects
- Built-in
- **GUI** & Command line
- **Merge conflicts**
- **Script** languages
- Extremely **flexible**
- Location of share
- **Packrat** and **Conda envs**

Easy setup

£0

RStudio

PyCharm

Learning

mergetools

Differencing

Branching

Best practice

Security

Online repos

Compatible



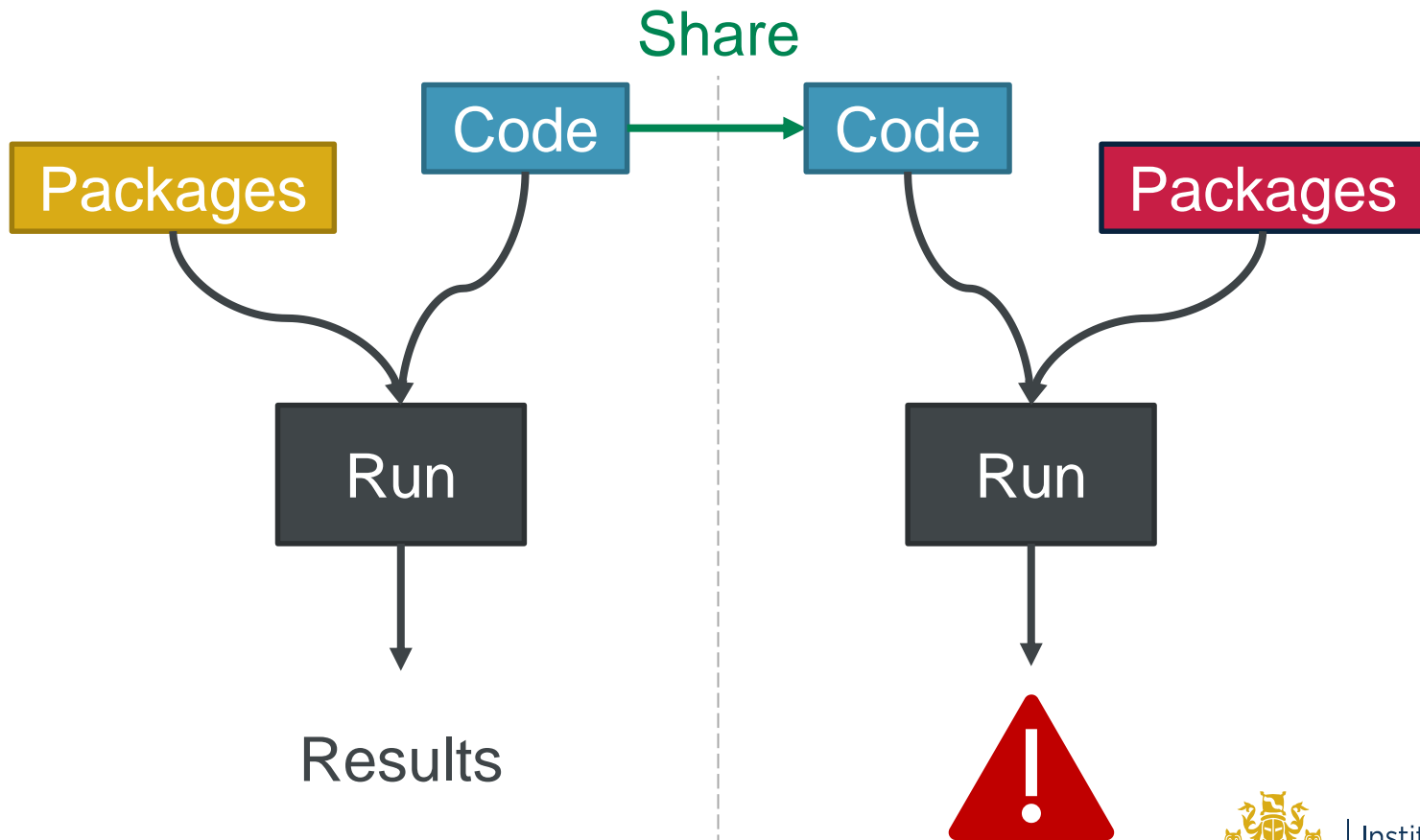
Institute  
and Faculty  
of Actuaries

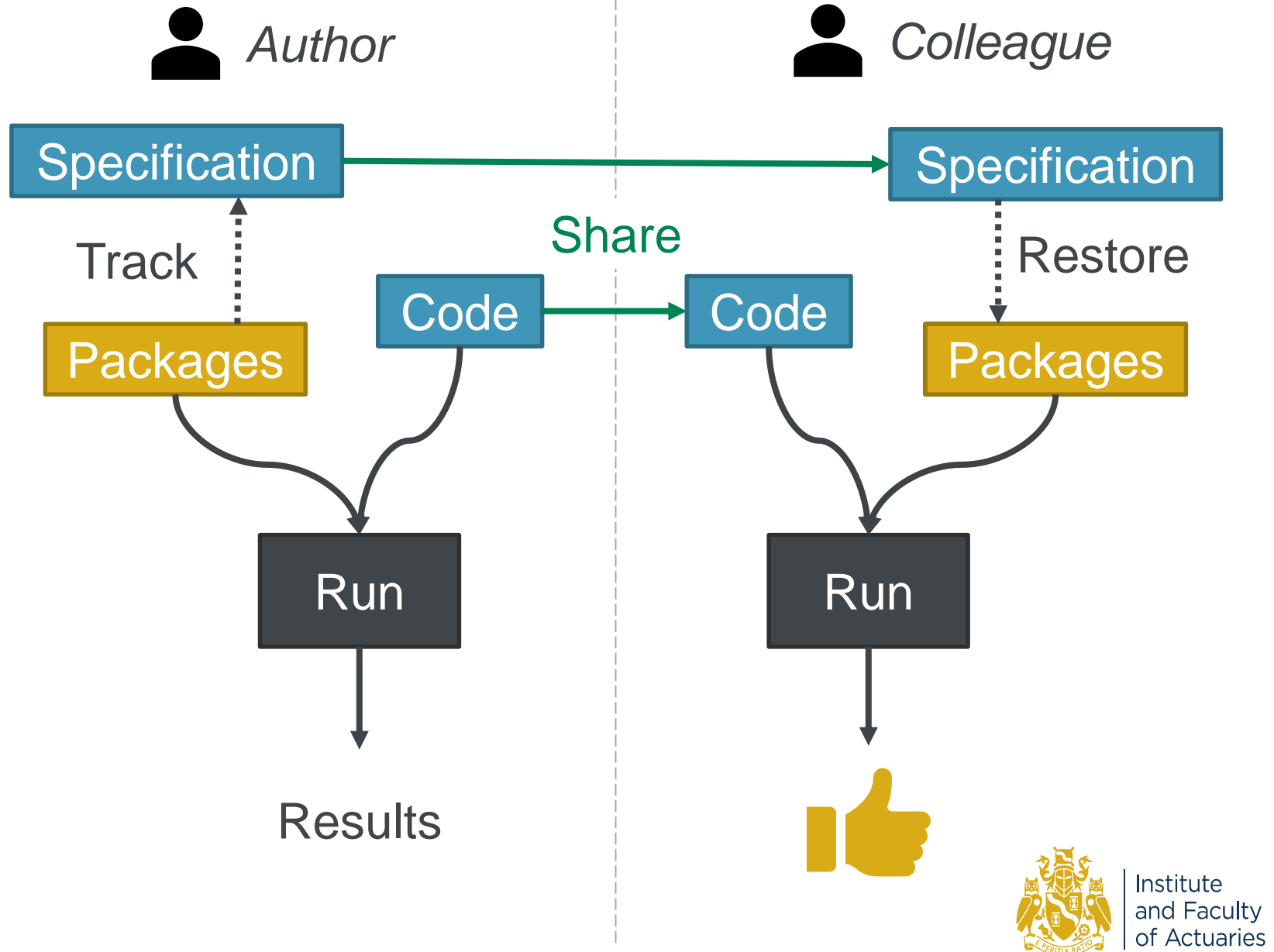


# Managing dependencies in R and Python



Institute  
and Faculty  
of Actuaries







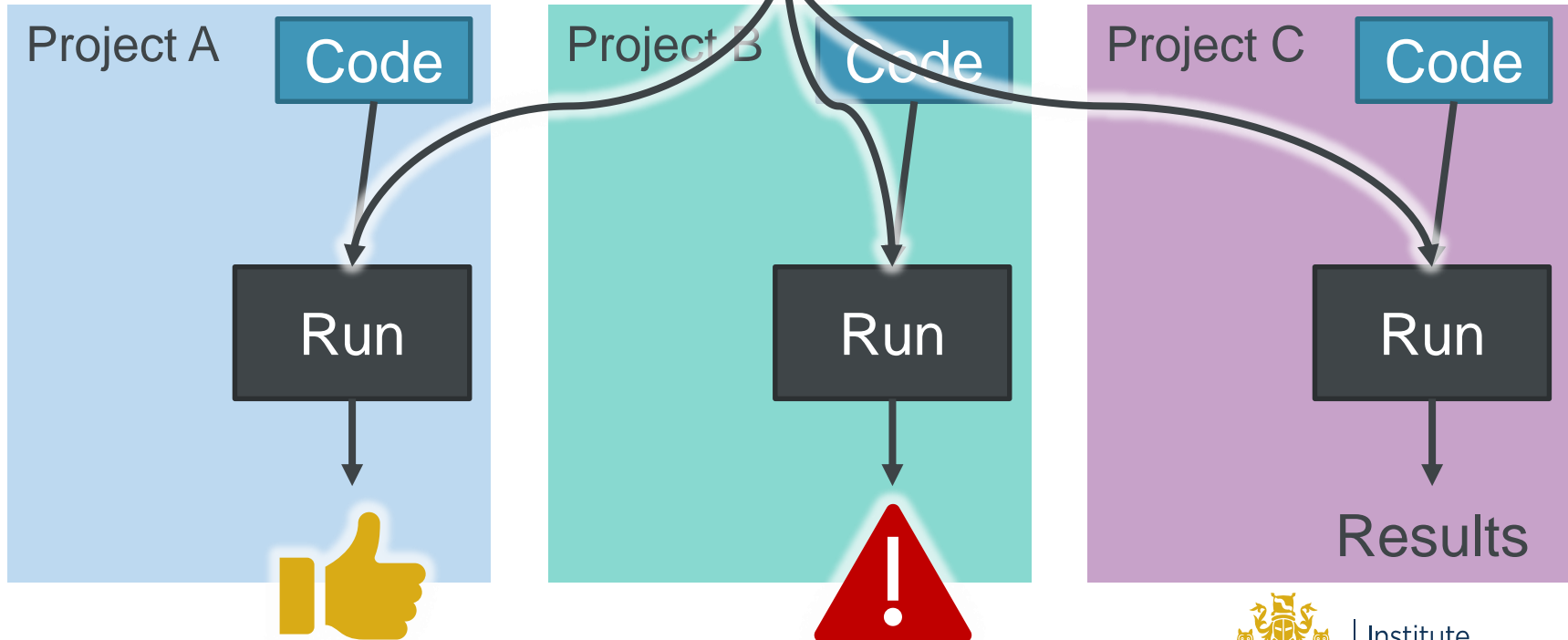
Internet

Download



Team member

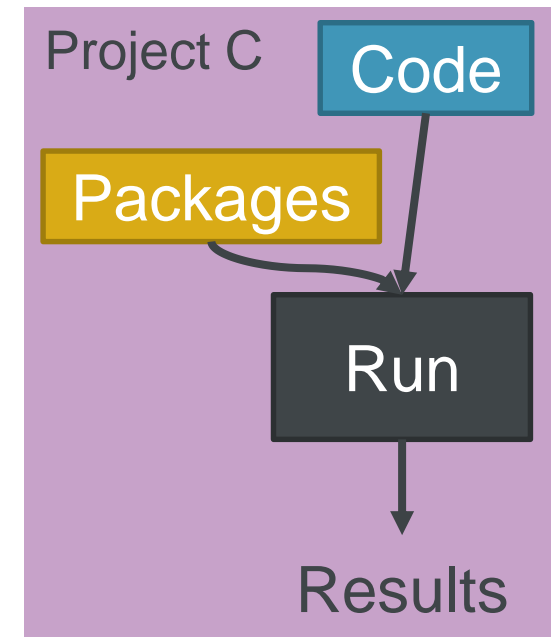
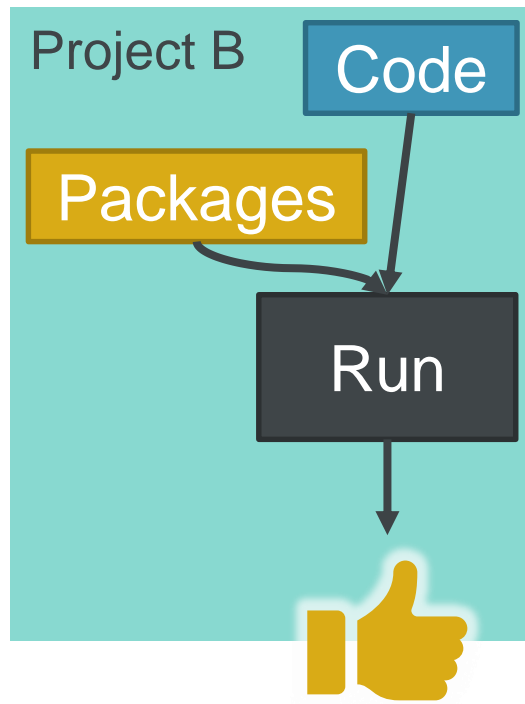
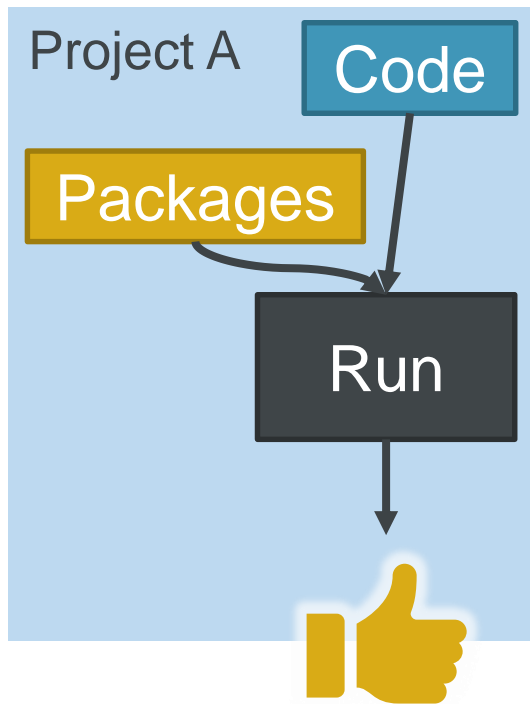
Personal Packages



Institute  
and Faculty  
of Actuaries

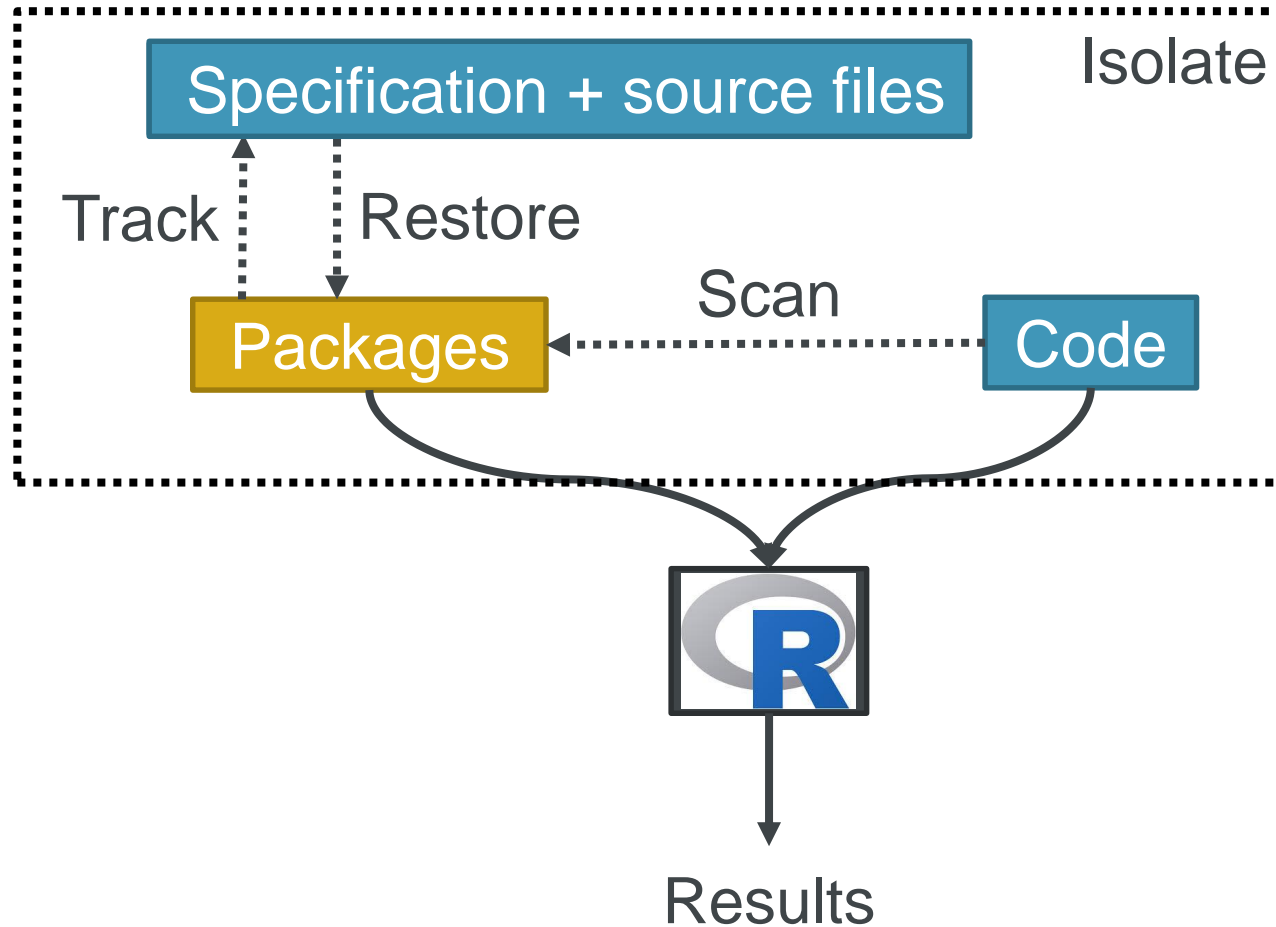


## Personal Packages



Institute  
and Faculty  
of Actuaries

# R + packrat approach





# Our findings: packrat

- Well **documented**
- **Integrates** with RStudio
- *Long* installation time ⇒  
Use the R GUI (not RStudio)
- Implement from the **start** of project
- Doesn't recognise all dependencies
- Only tracks packages, **not R itself**

Investigate *first*

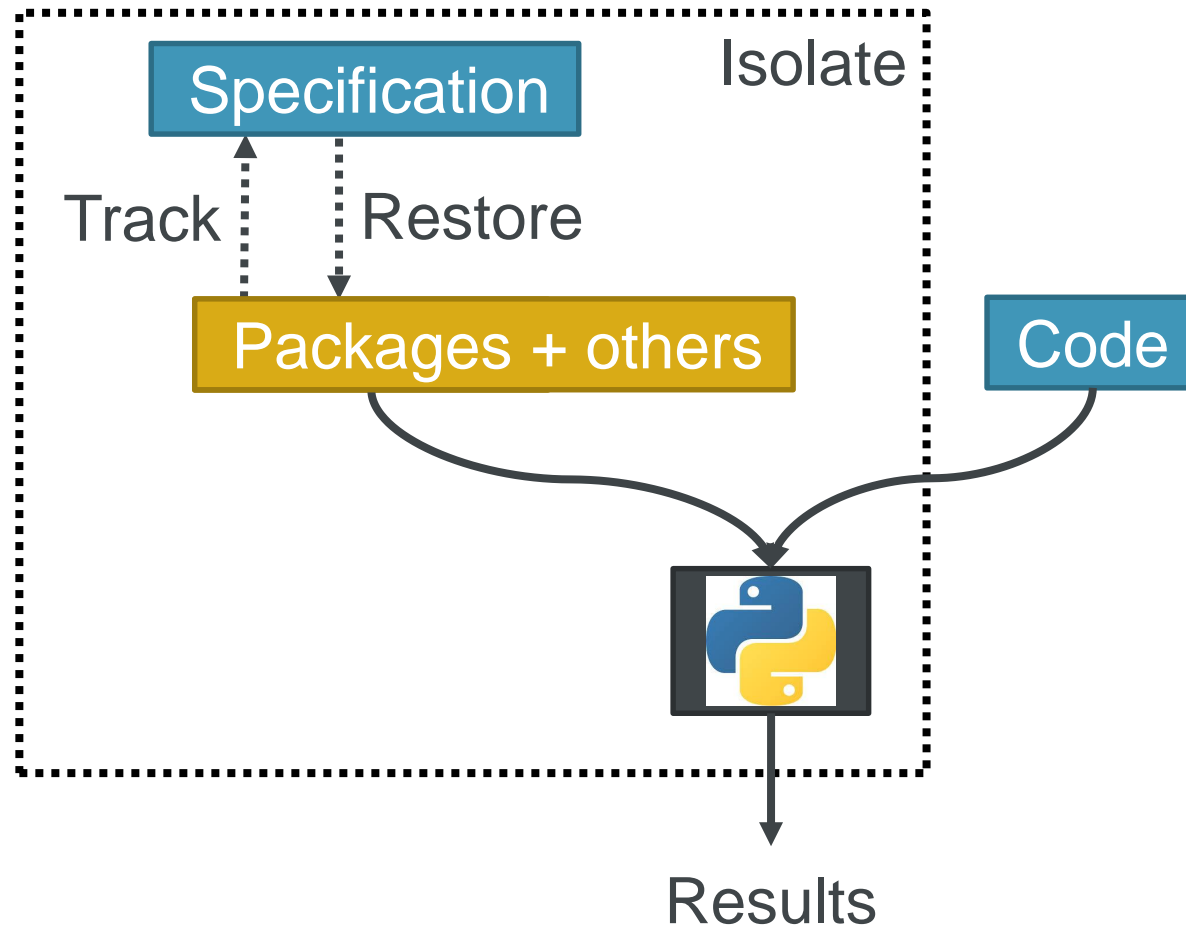
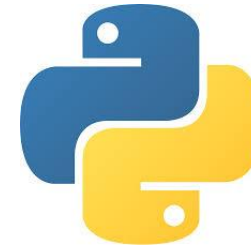
Add dummy script



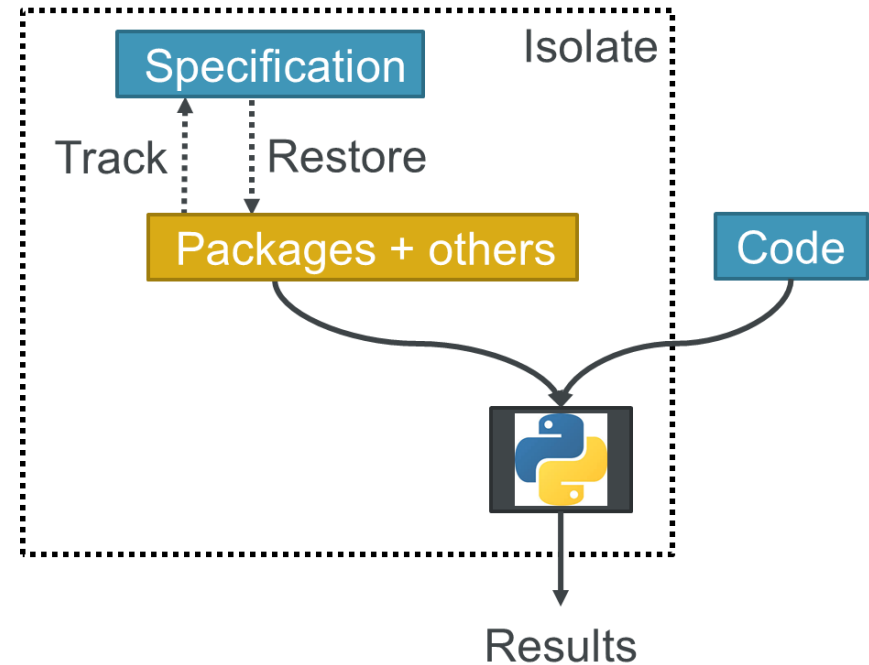
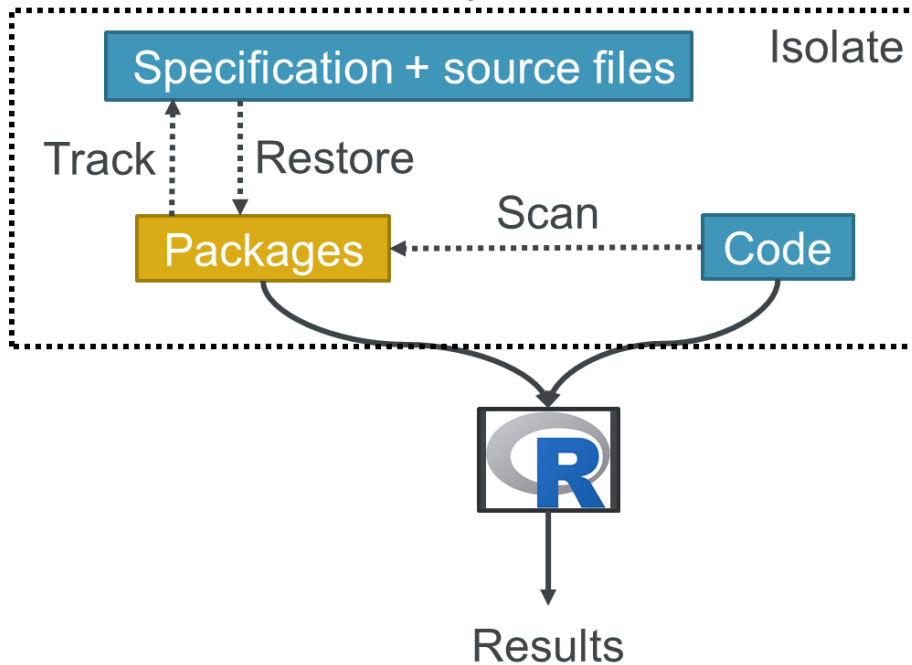
Institute  
and Faculty  
of Actuaries



# Python + conda approach



Institute  
and Faculty  
of Actuaries



*Not managing dependencies is not an option*



# Final thoughts



Institute  
and Faculty  
of Actuaries

# Data Science

Experiment

Learn

Share



Institute  
and Faculty  
of Actuaries

# Further resources

<b>Git</b>	Main website	<a href="https://git-scm.com/">https://git-scm.com/</a>
	Pro-git book (free)	<a href="https://git-scm.com/book/en/v2">https://git-scm.com/book/en/v2</a>
	Coursera course (free without certificate)	<a href="https://www.coursera.org/learn/version-control-with-git">https://www.coursera.org/learn/version-control-with-git</a>
	Git in Practice book (free)	<a href="https://github.com/GitInPractice/GitInPractice#readme">https://github.com/GitInPractice/GitInPractice#readme</a>
	DataCamp course: Intro to Git for Data Science (free)	<a href="https://www.datacamp.com/courses/introduction-to-git-for-data-science">https://www.datacamp.com/courses/introduction-to-git-for-data-science</a>
<b>packrat</b>	Main website	<a href="https://rstudio.github.io/packrat/">https://rstudio.github.io/packrat/</a>
	Using packrat with RStudio	<a href="https://www.rstudio.com/resources/webinars/managing-package-dependencies-in-r-with-packrat/">https://www.rstudio.com/resources/webinars/managing-package-dependencies-in-r-with-packrat/</a>
<b>conda</b>	Main website	<a href="https://conda.io/docs/index.html">https://conda.io/docs/index.html</a>
	Useful blog post “Conda: Myths and Misconceptions”	<a href="https://jakevdp.github.io/blog/2016/08/25/conda-myths-and-misconceptions/">https://jakevdp.github.io/blog/2016/08/25/conda-myths-and-misconceptions/</a>

Links checked at time of making this presentation in October 2018



Institute  
and Faculty  
of Actuaries

# Questions

# Comments

The views expressed in this [publication/presentation] are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this [publication/presentation] and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this [publication/presentation].

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this [publication/presentation] be reproduced without the written permission of the IFoA [or authors, in the case of non-IFoA research].



Institute  
and Faculty  
of Actuaries