University of Glasgow | College of Medical, Veterinary & Life Sciences

School of Biodiversity, One Health, and Veterinary Medicine

Key Research Skills II

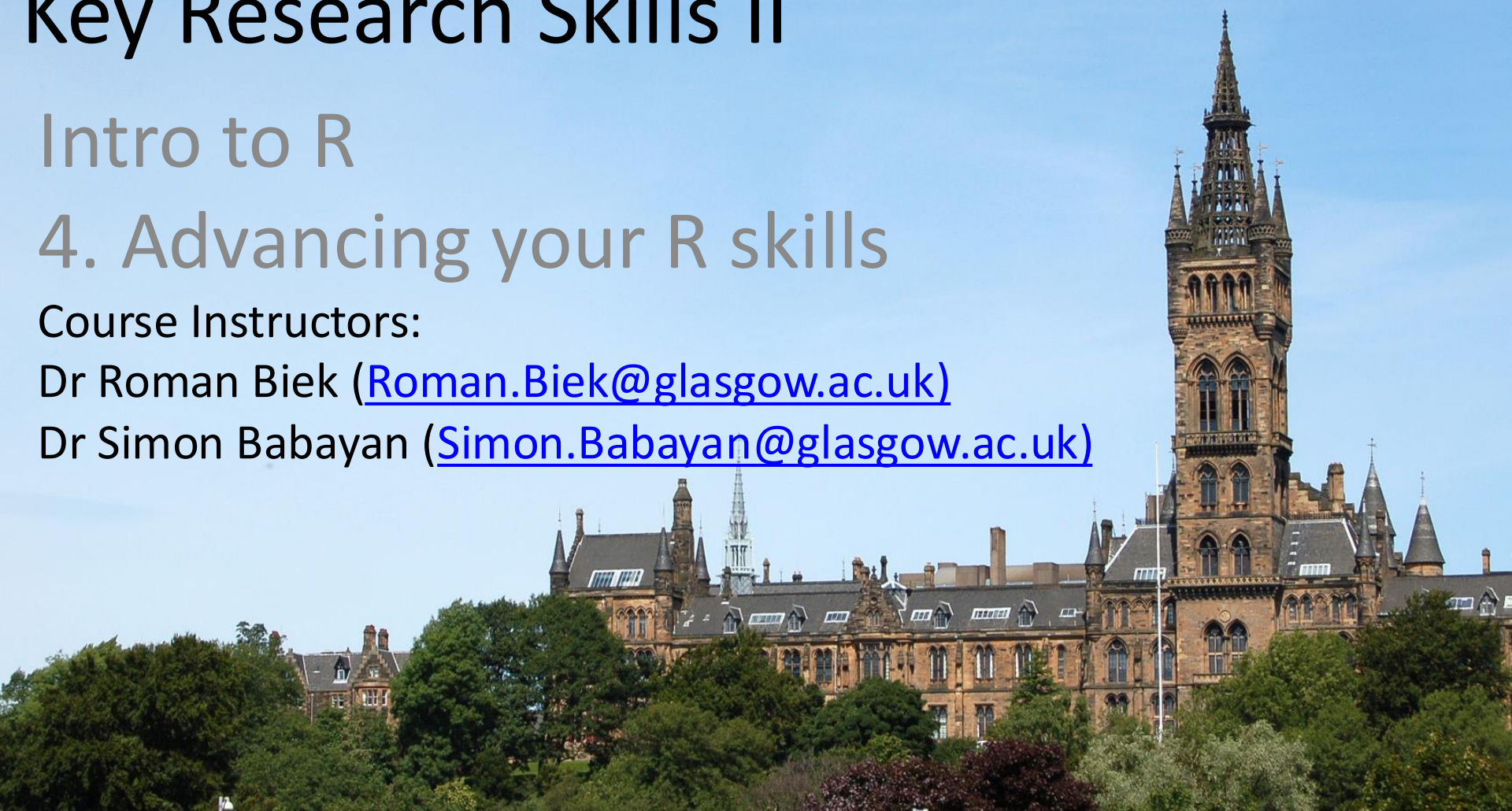
Intro to R

4. Advancing your R skills

Course Instructors:

Dr Roman Biek (Roman.Biek@glasgow.ac.uk)

Dr Simon Babayan (Simon.Babayan@glasgow.ac.uk)



Recap on notebooks & projects

- R Notebooks
  - Replaces R script, more advanced R markdown document
  - Easier to format and troubleshoot
- R projects
  - Easier to pick up where you left off
  - Keeps all files (e.g., data and code) associated in one place
- Why?
  - **Portability** (all in one project folder, easier to share)
  - **Reproducibility** (encourages clear annotation and structure of scripts and html reports)

Cleaning up scripts

- no need to explain obvious parts at this point (such as `rm(list=ls())` or `library()`)
- try to exclude tangents such as calling up help or calling up objects you've just created => use console for this, not script
- remove failed attempts (or 'comment out' using the # symbol)

Helping others

- Helping each other is encouraged
- But…
  - Try to provide pointers and explanations rather than simply sharing code
  - This allows others to learn from working through the problem instead of just copying solutions
  - It's not just about getting it right but understanding how to get it right so you can apply this understanding in the future

Errors in the R book

- List of identified errors and typos has been added to Moodle (next to link to book and datasets in the resources section of Day 1)
- If you find additional ones, contact us so they can be added to the list

Today's session:
Book chapter 5: statistics in R

- Book focusses on practical application, not statistical concepts => complementary to stats lectures
- If you feel unsure about the statistical background you should review materials for 'Intro to General Linear Models"
- Only will be doing Chapter 5 (Introducing Statistics) as part of this course – consider practicing Chapters 6 & 7 on your own

Todays session: Introducing statistics in R

- Chi-square contingency analysis
- Two sample t-test
- Simple linear regression
- One-way ANOVA

Start with those two

Complementary to what you learned in GLM sessions but more focus on using dplyr and ggplot2

Today's class

- **Start with section 5.4** (*Introducing … linear models*) before completing the earlier parts of this chapter
- Complete Chapter 5 in the book (all sections)
- Combine all sections into one script
- Watch out for some typos
  - in 5.3.1 use 'GardenOzone.csv' (not 'ozone.csv')

Challenge yourself…

- Try to replicate figures 5.11 (easier) and 5.10 (harder)
- Combines data points from multiple objects
- Also requires extra elements like
  - horizontal line: `geom_hline()`
  - line with specific start and end point: `geom_linerange()`

Additional take-home plotting exercise

- Intro to R: plotting
- Practice generating plots for a new dataset
- Available on Moodle by the end of the week

Assignments

- Finish today's exercise and upload script/notebook and HTML report by Friday 18th Oct, midnight [not graded; solutions will be provided]
- Complete take home exercise on plotting and upload script/notebook and HTML report by Thursday, 31st Oct midnight [graded]