Coding and Open Science Teaching in the Department of Earth Sciences

Lewis A. Jones

2026-02-25

Introduction

This report is a summary of the findings from the 2025 survey '*Coding and Open Science Teaching in the Department of Earth Sciences*'. The survey aimed to map current teaching practices around coding, computational methods, and open science across the Department of Earth Sciences at University College London. Survey findings are intended to guide a potential departmental-level strategy, and assess whether a dedicated 'Research Skills' module focused on coding and open science is needed.

Summary

This is the summary

Report

Below, a full report of the survey findings are given, organised around the survey questions. The report is generated in Quarto using R, with data processing and visualisaiton code chunks embedded for reproducibility purposes.

The source code and PDF version of the report can be downloaded here:

Download source | Download PDF

The raw and cleaned data from survey responses can be downloaded here:

Download raw data | Download clean data

Preamble

1. Load required libraries and data.

```
# Load required libraries
library(dplyr)
library(tibble)
library(tidyr)
library(readr)
library(stringr)
library(forcats)
library(ggplot2)
# Load data
dat <- read_csv("data/survey_responses.csv")
```

2. Clean and wrangle data.

```
# Some data cleaning and wrangling is required
dat <- dat %>%
  # Drop unneccesary columns
  select(-c(`Start time`, `Completion time`, `Last modified time`, Name, Email)) %>%
  # Rename columns for handling
  rename(Name = Name2,
         Email = Email2) %>%
  # Filter rows
  # Duplicate response: Pieter answered initially with a combined response for modules
  filter(ID != 2) %>%
  # Split Stephen's responses (multiple modules per response)
  bind_rows(., .[which(.$ID == 4), ]) %>%
  # Split Ana's responses (multiple modules per response)
  bind_rows(., .[which(.$ID == 21), ]) %>%
  # Split Alex's responses (multiple modules per response)
  bind_rows(., .[which(.$ID == 23), ]) %>%
  # Split Paul's responses (multiple modules per response)
  bind_rows(., .[which(.$ID == 24), ], .[which(.$ID == 24), ]) %>%
  # Sort by name
  arrange(Name) %>%
  # Add unique ID
  mutate(ID = 1:nrow(.)) %>%
  # Filter rows
  # Module reported twice
  # GEOL0044: Palaeoceanography
```

```r
  filter(ID != 6) %>%
  # GEOL0065: Research Methods and Skills
  filter(ID != 3) %>%
  # GEOL0018: Numerical Methods
  filter(ID != 4) %>%
  # Update unique ID
  mutate(ID = 1:nrow(.)) %>%
  # Mutate rows
  mutate(Email = tolower(Email),
         Name = str_to_title(Name),
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = case_when(
           `Module code and name (e.g. GEOL0017: Isotope Geology)` == "GEOL0012 Global Geophy
           `Module code and name (e.g. GEOL0017: Isotope Geology)` == "GEOL0004" ~ "GEOL0004
           `Module code and name (e.g. GEOL0017: Isotope Geology)` == "GEOL0045 Paleoclimate
           `Module code and name (e.g. GEOL0017: Isotope Geology)` == "GEOL0077" ~ "GEOL0077
           `Module code and name (e.g. GEOL0017: Isotope Geology)` == "GEOL0008 (Geochemistry
           `Module code and name (e.g. GEOL0017: Isotope Geology)` == "GEOL0069: AI4EO" ~ "GE
           ),
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 1, "GEOL003
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 2, "GEOL007
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 3, "GEOL003
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 12, "GEOL00
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 13, "GEOL00
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 14, "GEOL00
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 23, "GEOL00
         `Module code and name (e.g. GEOL0017: Isotope Geology)` = if_else(ID == 24, "GEOL00
         `Teaching level:` = if_else(ID == 1, "Undergraduate - Year 3", `Teaching level:`),
         `Teaching level:` = if_else(ID == 2, "Undergraduate - Year 2", `Teaching level:`),
         `Teaching level:` = if_else(ID == 3, "Undergraduate - Year 3", `Teaching level:`),
         `Teaching level:` = if_else(ID == 12, "Undergraduate - Year 1", `Teaching level:`),
         `Teaching level:` = if_else(ID == 13, "Undergraduate - Year 1", `Teaching level:`),
         `Teaching level:` = if_else(ID == 14, "Undergraduate - Year 4 (MSci/MSc)", `Teaching
         `Teaching level:` = if_else(ID == 23, "Undergraduate - Year 2", `Teaching level:`),
         `Teaching level:` = if_else(`Teaching level:` == "Undergraduate - Year 1;", "Undergr
         `Teaching level:` = if_else(`Teaching level:` == "Undergraduate - Year 2;", "Undergr
         `Teaching level:` = if_else(`Teaching level:` == "Undergraduate - Year 3;", "Undergr
         `Teaching level:` = if_else(`Teaching level:` == "Taught Masters (MSc);", "Taught Ma
         `Which programming languages or tools are used? (select all that apply)` = case_when
           `Which programming languages or tools are used? (select all that apply)` == "Pytho
           `Which programming languages or tools are used? (select all that apply)` == "R;QGI
           `Which programming languages or tools are used? (select all that apply)` == "MATLA
           `Which programming languages or tools are used? (select all that apply)` == "Not a
```

```
            `Which programming languages or tools are used? (select all that apply)` == "Excel
            `Which programming languages or tools are used? (select all that apply)` == "MATLA
            `Which programming languages or tools are used? (select all that apply)` == "Pytho
            `Which programming languages or tools are used? (select all that apply)` == "R;Exc
            `Which programming languages or tools are used? (select all that apply)` == "R;" ~
            `Which programming languages or tools are used? (select all that apply)` == "Pytho
            `Which programming languages or tools are used? (select all that apply)` == "MATLA
          ),
          `At what level are these coding skills taught?` = if_else(`At what level are these c
# Replace NA values
dat <- as_tibble(apply(dat, 2, function(x) { x[ is.na(x) ] <- "Not applicable"; x}))
write_csv(x = dat, file = "data/data.csv")
```

Which modules do we have data available for?

Thanks to everyone's responses, we have data for 24 modules from 17 different teaching staff. These responses covered the following modules:

```
# Print sorted unique list of modules
sort(unique(dat$`Module code and name (e.g. GEOL0017: Isotope Geology)`))
```

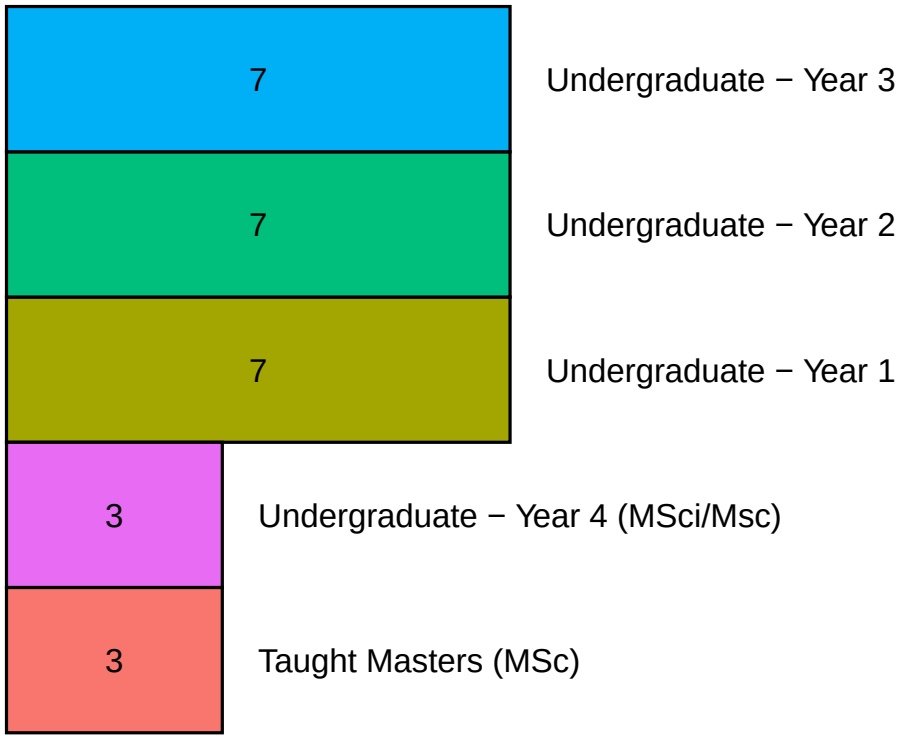
```
 [1] "GEOL0001: Earth Materials"
 [2] "GEOL0003: History of Life"
 [3] "GEOL0004: Dynamic Earth"
 [4] "GEOL0005: Foundations of Physical Geoscience"
 [5] "GEOL0007: The Earth"
 [6] "GEOL0008: Geochemistry"
 [7] "GEOL0009: Vertebrate Palaeontology and Evolution"
 [8] "GEOL0012: Global Geophysics"
 [9] "GEOL0013: Principles of Climate"
[10] "GEOL0015: Maps, Images & Structures"
[11] "GEOL0017: Isotope Geology"
[12] "GEOL0018: Numerical Methods"
[13] "GEOL0020/79: Field Geophysics"
[14] "GEOL0030: Seismology II"
[15] "GEOL0031: Environmental Geochemistry"
[16] "GEOL0036: Advanced Biodiversity and Macroevolutionary Studies"
[17] "GEOL0044: Palaeoceanography"
[18] "GEOL0045: Palaeoclimate and Palaeoenvironmental Change"
[19] "GEOL0057: Geodynamics and Global Tectonics"
[20] "GEOL0061: Statistics for Geoscientists"
```

```
[21] "GEOL0065: Research Methods and Skills"
[22] "GEOL0069: Artificial Intelligence for Earth Observations"
[23] "GEOL0076: Environmental Geoscience"
[24] "GEOL0077: Introduction to Mineral Resources"
```

What teaching levels do these modules cover?

Survey responses include modules from all teaching years. Note, several modules are available to students on different courses so the total exceeds 24.

```
# Define variable to visualise
var <- "Teaching level:"
# Summarise and plot
dat %>%
  select(.data[[var]]) %>%
  summarise(`Undergraduate - Year 1` = str_count(string = ., pattern = "Year 1"),
            `Undergraduate - Year 2` = str_count(string = ., pattern = "Year 2"),
            `Undergraduate - Year 3` = str_count(string = ., pattern = "Year 3"),
            `Undergraduate - Year 4 (MSci/Msc)` = str_count(string = ., pattern = "Year 4"),
            `Taught Masters (MSc)` = str_count(string = ., pattern = "Taught Masters")) %>%
  pivot_longer(cols = 1:5) %>%
  ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = name, x = value + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 15)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none")
```

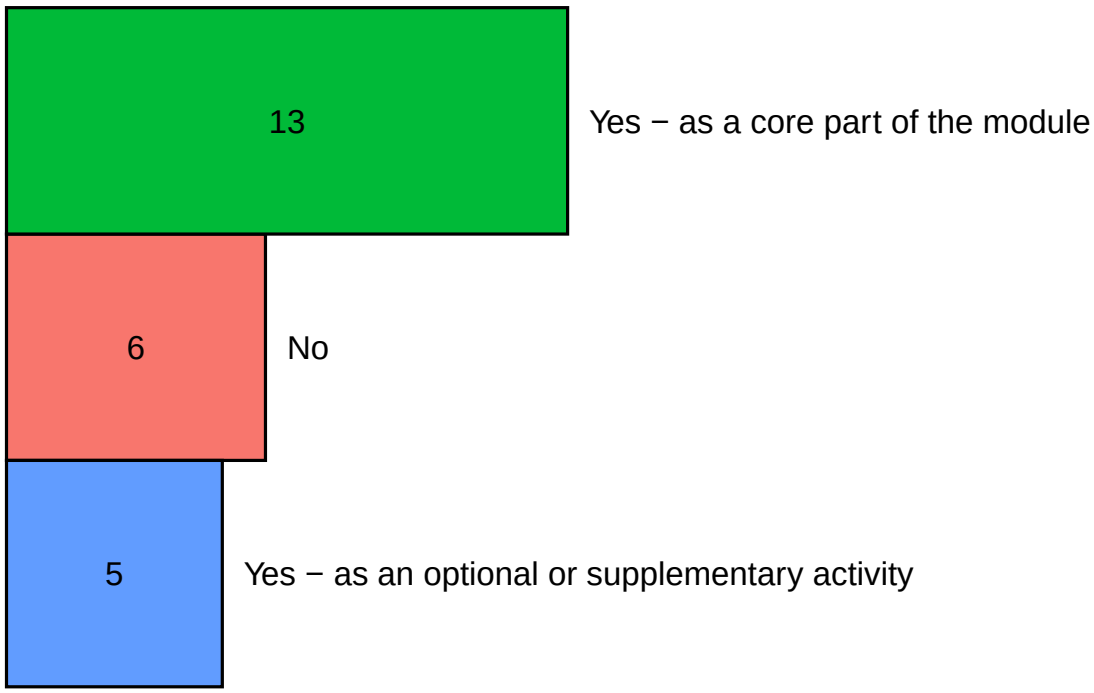


Are coding or computational elements included in the module?

Of the modules we have data for, coding or computational elements are currently a core component of ~54% of modules (13), and an optional or supplementary activity of ~21% of modules (5). The remaining 25% of modules (6) do not have any coding or computational element.

```
# Define variable for plotting
var <- "Do you include coding or computational elements in the module?"
# Summarise and visualise data
dat %>%
  group_by(.data[[var]]) %>%
  summarise(n = length(.data[[var]])) %>%
  ggplot(data = ., aes(x = n, y = fct_reorder(.data[[var]], n), fill = .data[[var]])) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = .data[[var]], x = n + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = n), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 25)) +
    xlab("Counts") +
    theme_void() +
```

```
    theme(legend.position = "none",
          axis.title.x = element_text())
```



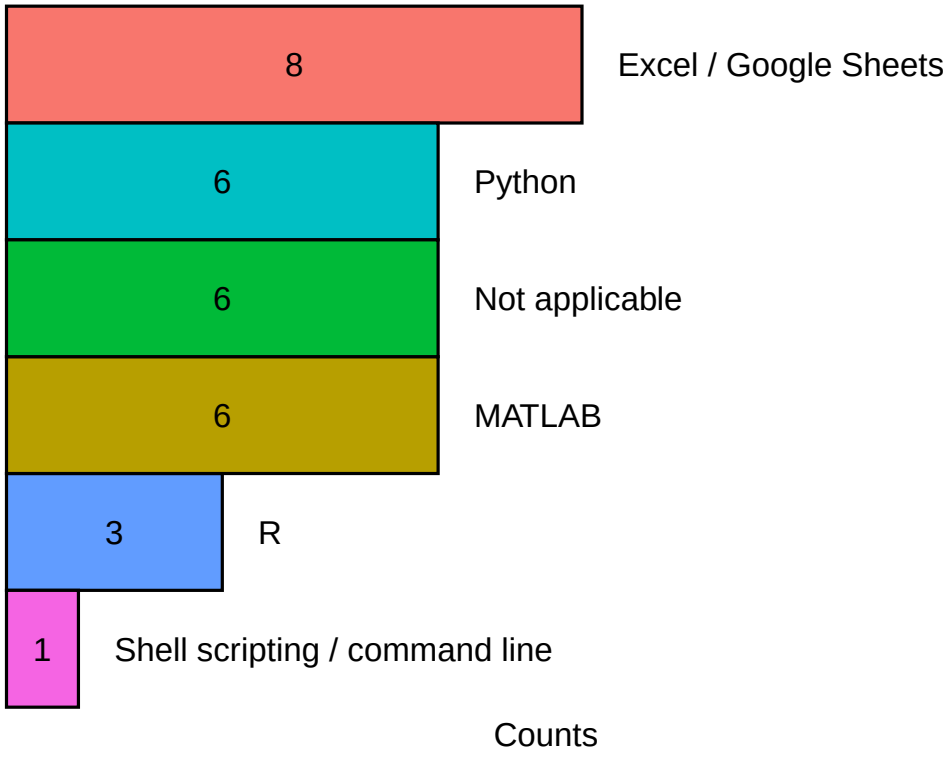
Which programming languages or tools are used?

Across modules, students are currently exposed to the following programming languages or tools: Excel / Google Sheets (~33% of modules), Python (~25% of modules), MATLAB (~25% of modules), R (~13% of modules), Shell scripting / command line (~4% of modules). In six modules (~25%), students are not currently exposed to any programming languages or computational tools. Note, the sum of percentages is greater than 100% as some modules contain multiple languages/tools.

```
var <- "Which programming languages or tools are used? (select all that apply)"

dat %>%
  select(.data[[var]]) %>%
  summarise(R = str_count(string = ., pattern = "R"),
            MATLAB = str_count(string = ., pattern = "MATLAB"),
            `Shell scripting / command line` = str_count(string = ., pattern = "Shell scripti
```

```
          `Excel / Google Sheets` = str_count(string = ., pattern = "Excel / Google Sheets"
          Python = str_count(string = ., pattern = "Python"),
          `Not applicable` = str_count(string = ., pattern = "Not applicable")) %>%
pivot_longer(cols = 1:6) %>%
ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
  geom_col(width = 1, colour = "black") +
  geom_text(aes(label = name, x = value + 0.5), hjust = 0, colour = "black") +
  geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
  scale_x_continuous(limits = c(0, 15)) +
  xlab("Counts") +
  theme_void() +
  theme(legend.position = "none",
        axis.title.x = element_text())
```



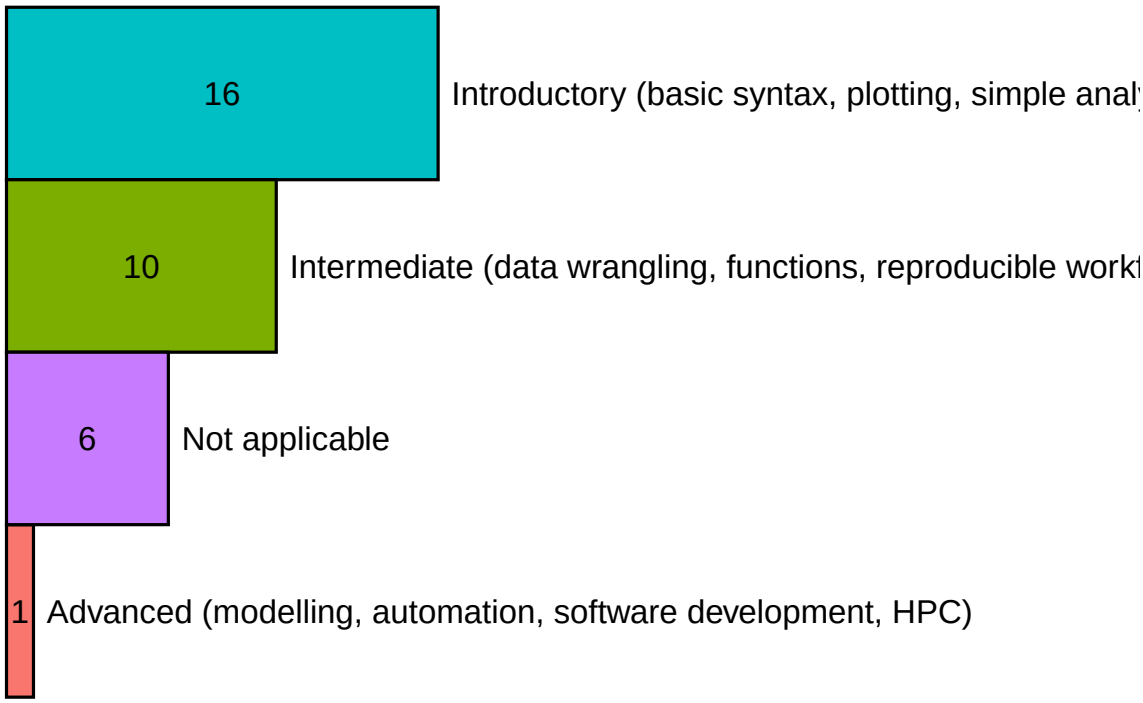
At what level are these coding skills taught?

Across the 24 modules recorded, students are exposed to an introductory (~66% of modules), intermediate (~41% of modules), and advanced (~4%) level of coding skills. As above, in six

modules (~25%), students are not currently exposed to coding. Note, the sum of percentages is greater than 100% as some modules contain multiple languages/tools.

```
var <- "At what level are these coding skills taught?"

dat %>%
  select(.data[[var]]) %>%
  summarise(`Introductory (basic syntax, plotting, simple analysis)` = str_count(
              string = ., pattern = "Introductory"
              ),
            `Intermediate (data wrangling, functions, reproducible workflows)` = str_count(
              string = ., pattern = "Intermediate"
              ),
            `Advanced (modelling, automation, software development, HPC)` = str_count(
              string = ., pattern = "Advanced"
              ),
            `Not applicable` = str_count(string = ., pattern = "Not applicable")
            ) %>%
  pivot_longer(cols = 1:4) %>%
  ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = name, x = value + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 40)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```



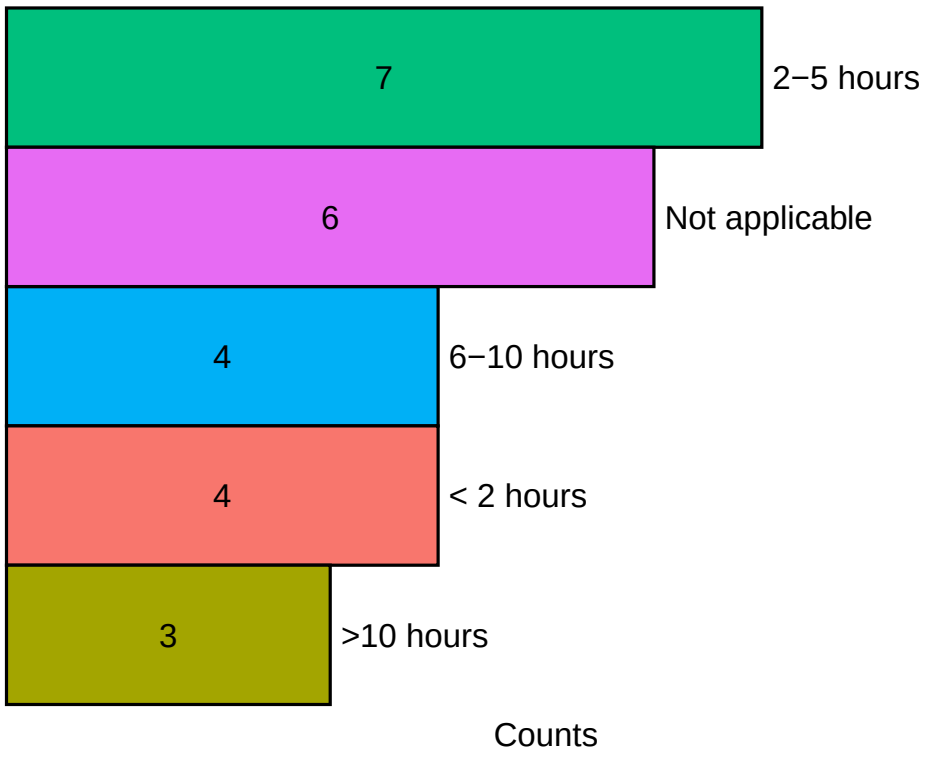
How much time (in contact hours) is dedicated to coding or computational training within the module?

The contact hours dedicated to coding or computational training is variable across modules, with ~13% of modules (n = 3) exposing students to more than 10 dedicated hours of coding or computational training, while ~42% of modules have less than 2 hours or no exposure (n = 10).

```
var <- "How much time (in contact hours) is dedicated to coding or computational training wi

dat %>%
  group_by(.data[[var]]) %>%
  summarise(n = length(.data[[var]])) %>%
  ggplot(data = ., aes(x = n, y = fct_reorder(.data[[var]], n), fill = .data[[var]])) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = .data[[var]], x = n + 0.1), hjust = 0, colour = "black", ) +
    geom_text(aes(label = n), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 10)) +
    xlab("Counts") +
    theme_void() +
```

```
    theme(legend.position = "none",
          axis.title.x = element_text())
```



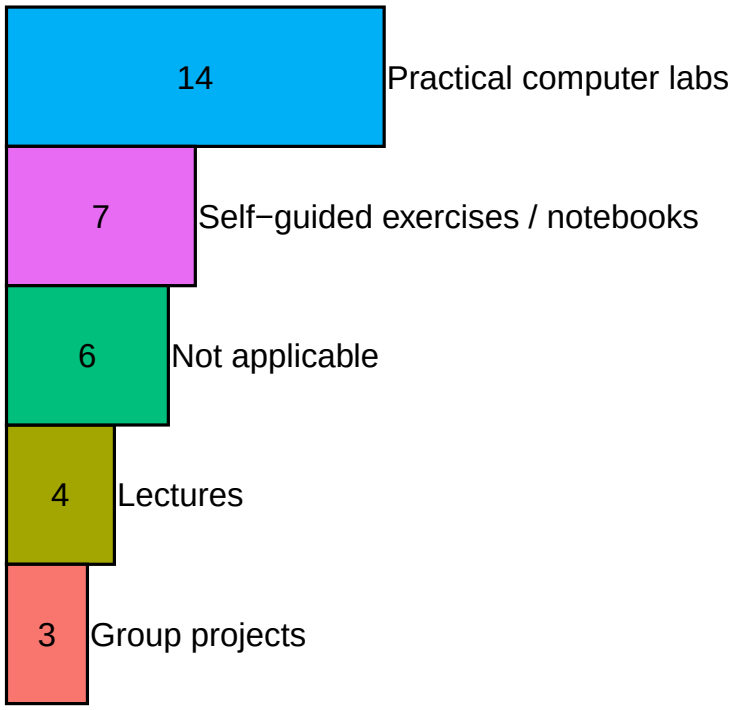
How are these skills taught?

A range of approaches are used to deliver computation training with practical computer labs being the most common approach (~58% of modules). Perhaps unsurprisingly, lecture and group projects are the least common delivery approach for computational training.

```
var <- "How do you teach these skills? (select all that apply)"

dat %>%
  select(.data[[var]]) %>%
  summarise(`Practical computer labs` = str_count(
              string = ., pattern = "computer labs|practical sessions"
              ),
            `Self-guided exercises / notebooks` = str_count(
              string = ., pattern = "notebooks"
              ),
```

```
            `Lectures` = str_count(
              string = ., pattern = "Lectures"
              ),
            `Group projects` = str_count(
              string = ., pattern = "Group projects"
              ),
            `Not applicable` = str_count(string = ., pattern = "Not applicable")
            ) %>%
  pivot_longer(cols = 1:5) %>%
  ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = name, x = value + 0.1), hjust = 0, colour = "black") +
    geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 40)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```



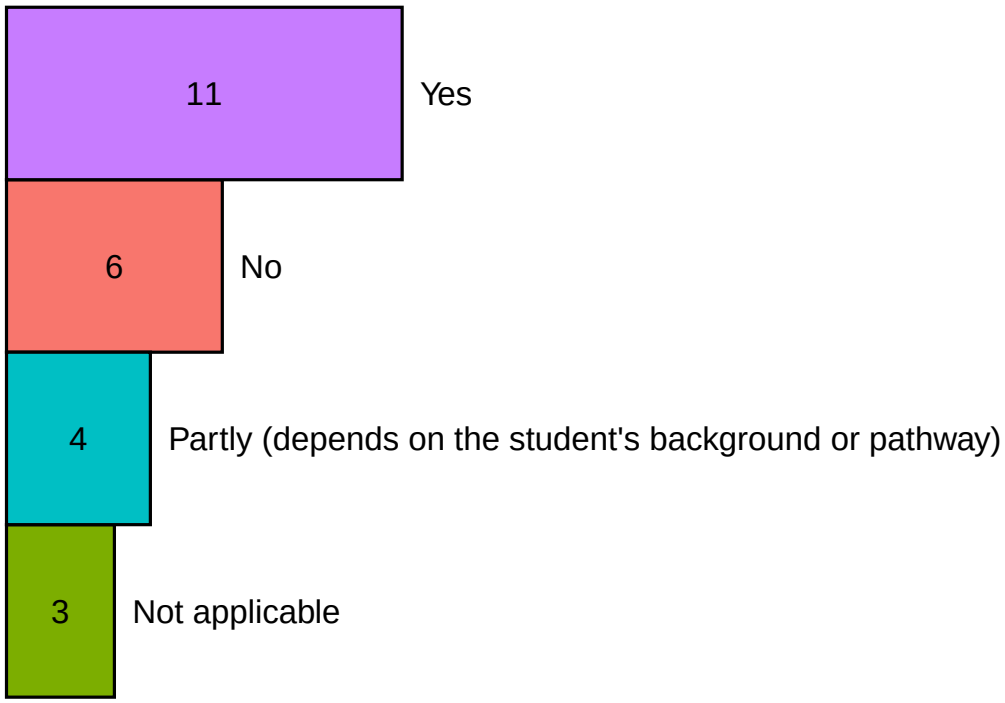
How do you assess coding or computational skills in the module?

Text needed

Do you assume that students taking your module already have prior coding experience or knowledge?

```
var <- "Do you assume that students taking your module already have prior coding experience o

dat %>%
  group_by(.data[[var]]) %>%
  summarise(n = length(.data[[var]])) %>%
  filter(!is.na(.data[[var]])) %>%
  ggplot(data = ., aes(x = n, y = fct_reorder(.data[[var]], n), fill = .data[[var]])) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = .data[[var]], x = n + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = n), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 30)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```



If yes or partly, please specify:Which previous module(s) or experience do you assumeWhat level of coding competency you expect (e.g. basic syntax, data plotting, use of scripts)

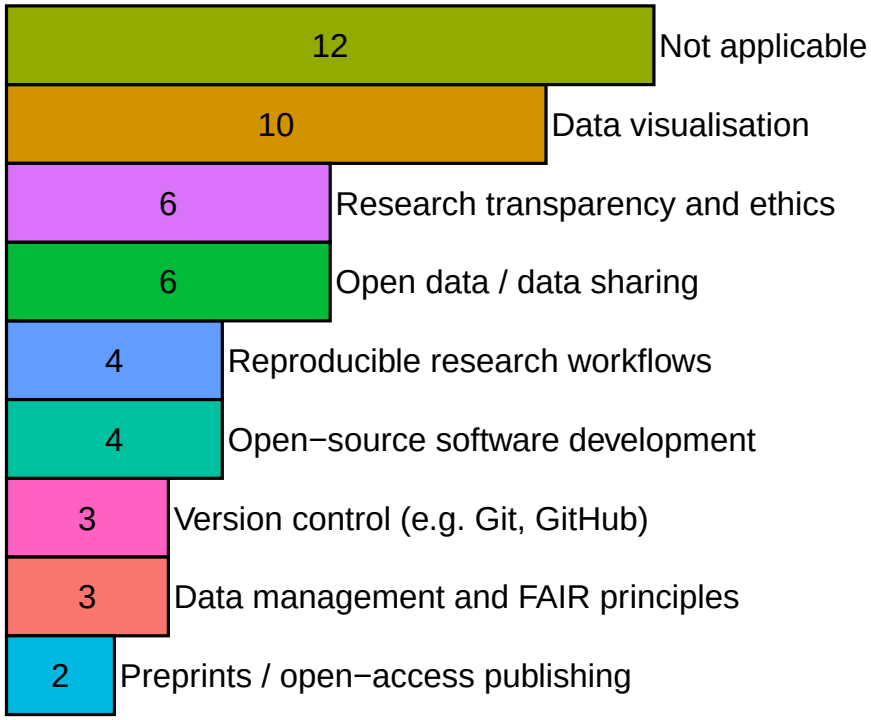
Text response

Do you explicitly teach or discuss any of the following open science practices in the module? (select all that apply)

```
var <- "Do you explicitly teach or discuss any of the following open science practices in the

dat %>%
  select(.data[[var]]) %>%
  summarise(`Version control (e.g. Git, GitHub)` = str_count(
              string = ., pattern = "Version control"
              ),
            `Data management and FAIR principles` = str_count(
              string = ., pattern = "Data management and FAIR principles"
```

```
            ),
          `Reproducible research workflows` = str_count(
            string = ., pattern = "Reproducible research workflows"
            ),
          `Data visualisation` = str_count(
            string = ., pattern = "Data visualisation"
            ),
          `Open data / data sharing` = str_count(
            string = ., pattern = "Open data / data sharing"
            ),
          `Open-source software development` = str_count(
            string = ., pattern = "Open-source software development"
            ),
          `Preprints / open-access publishing` = str_count(
            string = ., pattern = "Preprints / open-access publishing"
            ),
          `Research transparency and ethics` = str_count(
            string = ., pattern = "Research transparency and ethics"
            ),
          `Not applicable` = str_count(string = ., pattern = "Not applicable")
          ) %>%
pivot_longer(cols = 1:9) %>%
ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
  geom_col(width = 1, colour = "black") +
  geom_text(aes(label = name, x = value + 0.1), hjust = 0, colour = "black") +
  geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
  scale_x_continuous(limits = c(0, 20)) +
  xlab("Counts") +
  theme_void() +
  theme(legend.position = "none",
        axis.title.x = element_text())
```



If taught, how are these topics integrated into the module?

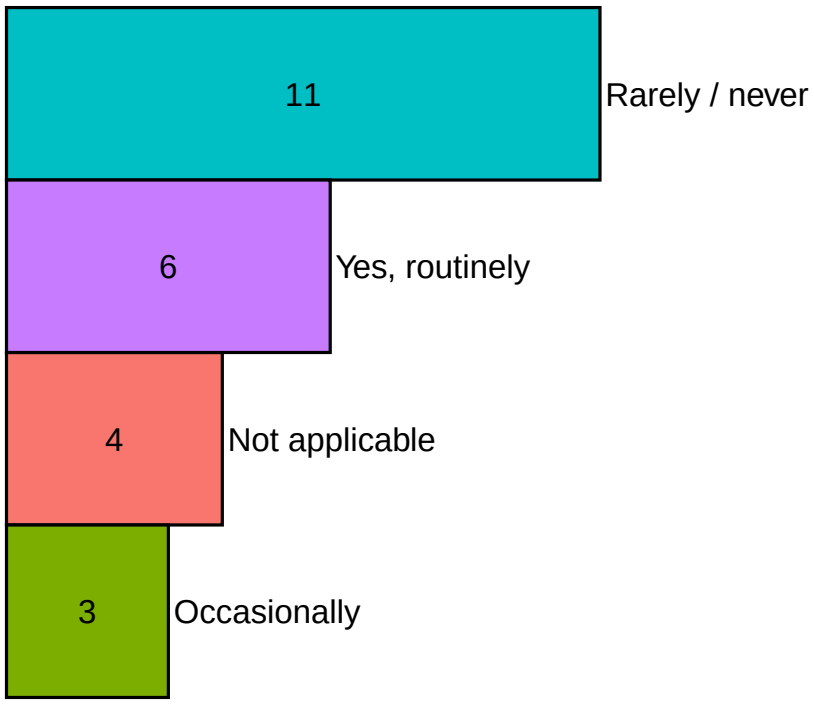
Text response needed

In the module, do students produce or use openly available code or datasets as part of their coursework or projects?

```
var <- "In the module, do students produce or use openly available code or datasets as part o

dat %>%
  group_by(.data[[var]]) %>%
  summarise(n = length(.data[[var]])) %>%
  filter(!is.na(.data[[var]])) %>%
  ggplot(data = ., aes(x = n, y = fct_reorder(.data[[var]], n), fill = .data[[var]])) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = .data[[var]], x = n + 0.1), hjust = 0, colour = "black") +
    geom_text(aes(label = n), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 20)) +
```

```
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```

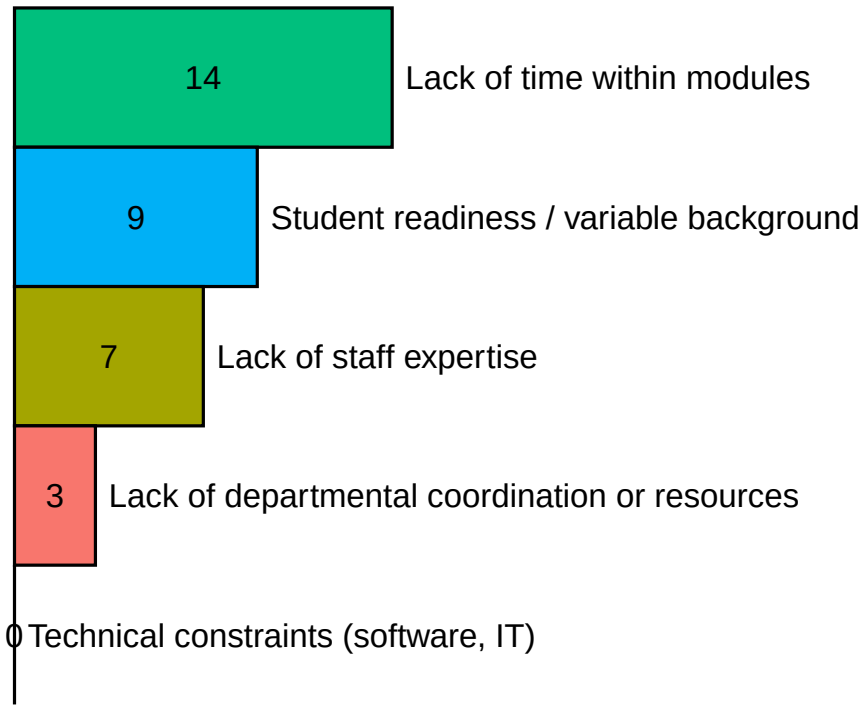


What do you think are the main barriers to teaching open science practices? (select all that apply)

```
var <- "What do you think are the main barriers to teaching open science practices? (select a

dat %>%
  select(.data[[var]]) %>%
  summarise(`Lack of time within modules` = str_count(
              string = ., pattern = "Lack of time within modules"
              ),
            `Student readiness / variable background` = str_count(
              string = ., pattern = "Student readiness"
              ),
```

```
            `Lack of staff expertise` = str_count(
              string = ., pattern = "Lack of staff expertise"
              ),
            `Lack of departmental coordination or resources` = str_count(
              string = ., pattern = "Lack of departmental coordinates or resources"
              ),
           `Technical constraints (software, IT)` = str_count(
              string = ., pattern = "Technical constraints"
              ),
            `Not applicable` = str_count(string = ., pattern = "Not applicable")
           ) %>%
  pivot_longer(cols = 1:5) %>%
  ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = name, x = value + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 40)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```

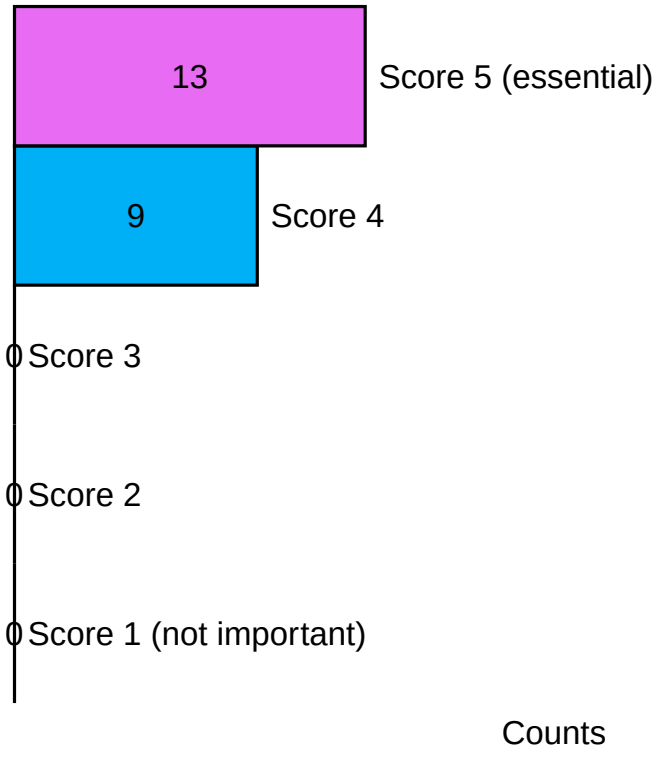


How important do you think coding and open science skills are for Earth Science graduates? (1 = not important, 5 = essential)

```
var <- "How important do you think coding and open science skills are for Earth Science gradu

dat %>%
  select(.data[[var]]) %>%
  summarise(`Score 1 (not important)` = str_count(
              string = ., pattern = "1"
              ),
            `Score 2` = str_count(
              string = ., pattern = "2"
              ),
            `Score 3` = str_count(
              string = ., pattern = "3"
              ),
            `Score 4` = str_count(
              string = ., pattern = "4"
              ),
```

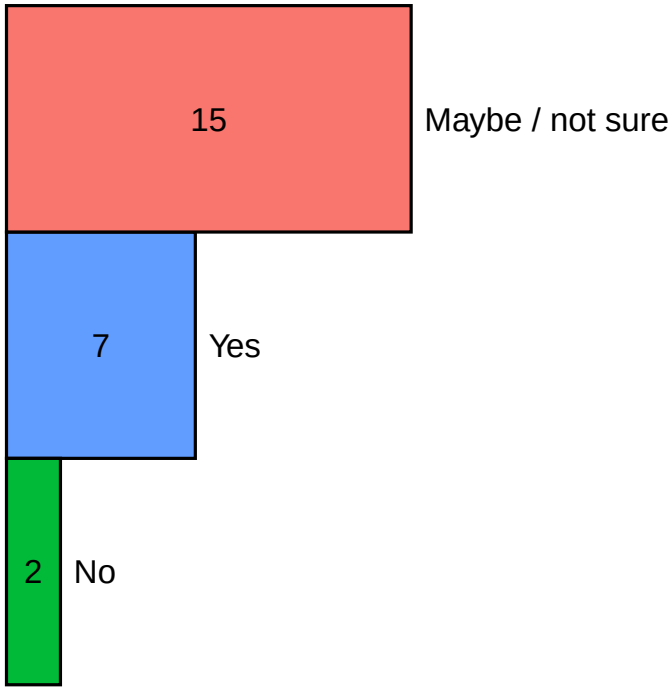
```
            `Score 5 (essential)` = str_count(
              string = ., pattern = "5"
              )
            ) %>%
  pivot_longer(cols = 1:5) %>%
  ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = name, x = value + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 40)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```



Do you think there is a need for a dedicated departmental module on coding and open science skills?

```
var <- "Do you think there is a need for a dedicated departmental module on coding and open s

dat %>%
  select(.data[[var]]) %>%
  summarise(`Yes` = str_count(
              string = ., pattern = "Yes"
              ),
            `Maybe / not sure` = str_count(
              string = ., pattern = "Maybe / not sure"
              ),
            `No` = str_count(
              string = ., pattern = "No"
              )
            ) %>%
  pivot_longer(cols = 1:3) %>%
  ggplot(data = ., aes(x = value, y = fct_reorder(name, value), fill = name)) +
    geom_col(width = 1, colour = "black") +
    geom_text(aes(label = name, x = value + 0.5), hjust = 0, colour = "black") +
    geom_text(aes(label = value), position = position_stack(vjust = 0.5), colour = "black") +
    scale_x_continuous(limits = c(0, 40)) +
    xlab("Counts") +
    theme_void() +
    theme(legend.position = "none",
          axis.title.x = element_text())
```



If yes or maybe, what do you think such a module should include?

Text

What support or resources would help you integrate more coding or open science training into your teaching?

Text

Any additional comments or suggestions?

Text