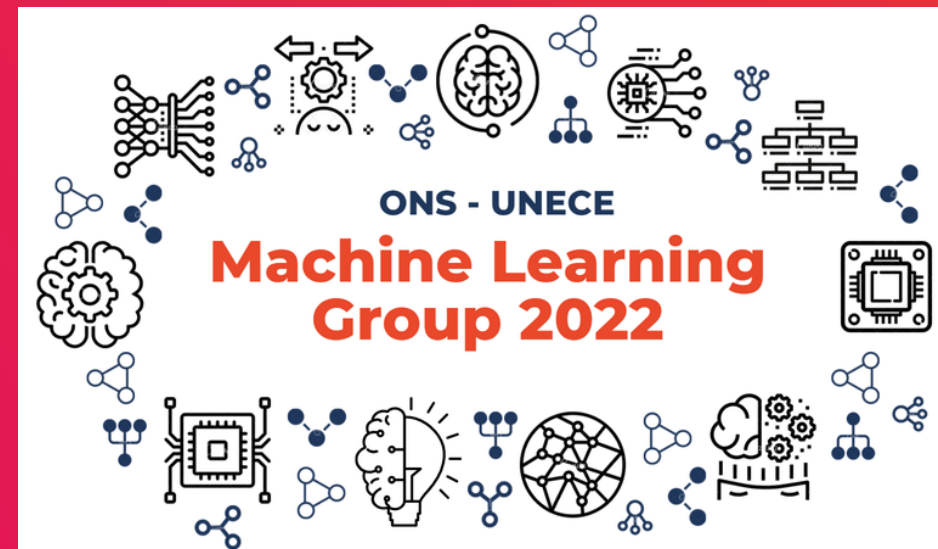


# Coffee and Coding Welcome



## ONS-UNECE Machine Learning Group 2022

2 November 2022

# Housekeeping



## Recording

Today's webinar will be recorded. The recording and slides will be available on the UNECE website after the event.



At the end of the seminar, you will be asked to take part in a short survey.



## Questions and comments

The Teams chat will be closed during the presentation and will reopen at the end of the presentation.

Please use Slido to take part in polls and post your questions:

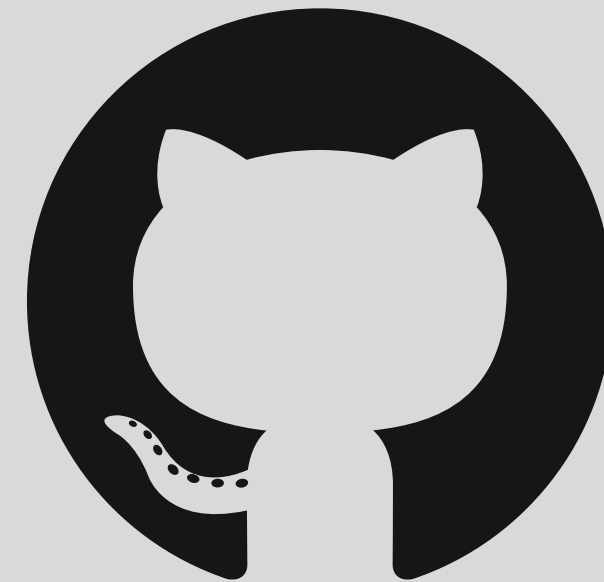
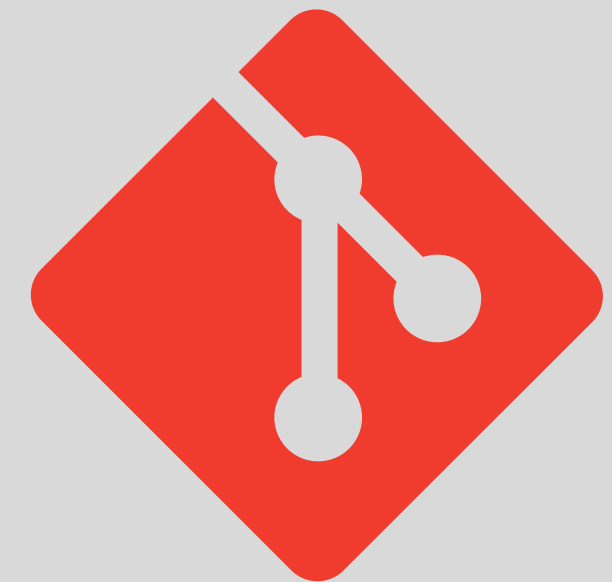
Go to  
[www.slido.com](https://www.slido.com)

Slido Code:  
#3881745



# Git Good

A quick introduction to Git and  
Github

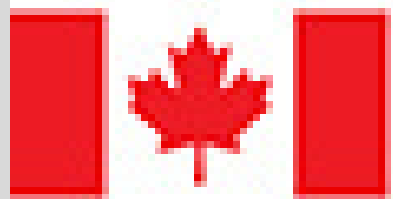


# Questions?

Ask in the chat! Brittany or I will see them!

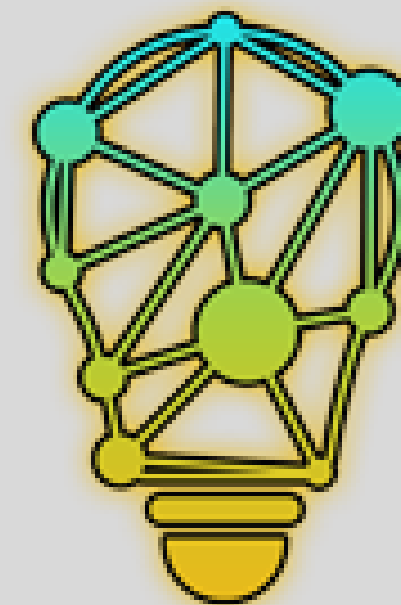


# Who are we?



Statistics  
Canada

Statistique  
Canada



# Agenda



- 1 Introduction
- 2 Preconceived Notions
- 3 What are Git and GitHub?
- 4 Break
- 5 GitHub Demo/Follow Along
- 6 Close-Out



# Preconceived Notions

Head to the slido link in the chat  
You have 2 minutes!

# Why version controlling code matters for NSOs

Reproducibility

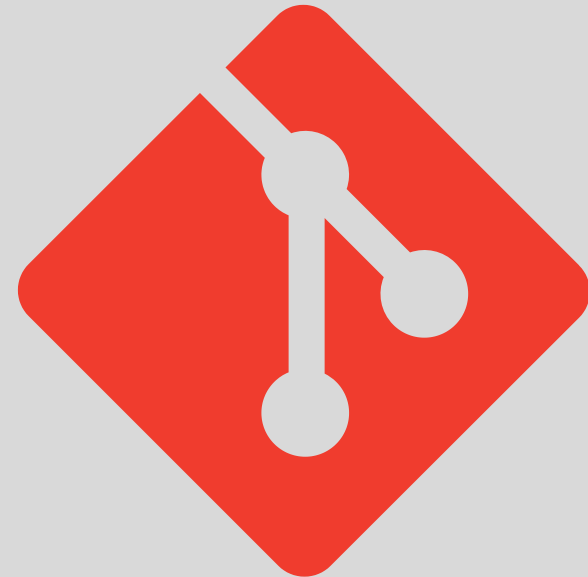
Efficient and effective development of code

Create standardization between and within organizations

Easily disseminate changes and updated versions of programs



# What is Git?



- Git is an open source distributed version control system
- It is installed locally on your computer
- We are familiar with other forms of version control such as:
  - Microsoft Word's track changes
  - Google Docs' version history

# What is GitHub/GitLab

- GitHub and GitLab are web-based services that allow for collaborative coding and dissemination of code
- They can be installed on internal servers too!
- Both GitHub and GitLab have additional project management and collaboration features to help with workflows
- You can select who sees what, so not everything has to be open to everyone if you don't want it to

- Git connects to GitHub, so we make our changes then use Git to upload them to GitHub
- Multiple people can connect their Git to a GitHub project to collaborate, and new collaborators can be added at any time

Share changes and get updates with GitHub

push changes



pull updates



Work locally with Git

# Answers to some of your questions

Yes, you can use GitHub with RStudio and VSCode (they both have a tutorial on it)

GitHub vs GitLab, it is all in the CI

Keep your data off of GitHub with .gitignore files and Git Status

Pull updates frequently to avoid merge conflicts



Break!



10 Minute break

Log into Github.com when you come  
back

# Useful resources

Pages / Machine Learning for Official Statistics Home

## Machine Learning Group 2022

Created by Inkyung Choi, last modified by Alison Baily on 27 Oct, 2022

### Training Event

#### Invitation to join Coffee and Coding session on Git

The last of this year's ONS - UNECE Machine Learning Groups Coffee and Coding session is on 2 November 2022 at 1400 - 1530 (CEST) / 0900 - 1030 (EST). (Time zone converter here).

In this session, Tabitha Williams and Britny Vongdara from Statistics Canada will provide an interactive lesson on using GitHub, and an introduction to Git. They'll go through forking a repository, making a commit, collaboration, and how to avoid uploading your data on GitHub. The second half of the session will be theory and a discussion on the difference between GitHub and Git, what a Git project looks like normally, and best practices.

For more information and to register, please visit the following Eventbrite page:

- Coffee and Coding Session 2 November

For any questions, please don't hesitate to get in touch with [ML2022@ons.gov.uk](mailto:ML2022@ons.gov.uk)

### International collaboration on Machine Learning in Official Statistics

We are an international platform for research collaboration, knowledge exchange and capability building on ML for official statistics. We bring together 420+ members from 44 different countries to explore how ML can improve statistical output and be integrated successfully into production. Last year we ran 18 different research projects and held regular expert presentations on ML innovations at statistical organisations around the world.

We're a dynamic and growing community, and welcome all levels of ML experience from expert to novice. It's a great place to connect with and learn from others working on ML. We're especially keen to recruit more experienced data scientists to our membership this year.

This year our programme focuses on the following themes:

#### Applications

- Big Data
- Coding and Classification
- Modelling
- Imagery

#### Implementation

- Quality of training data and algorithms
- Model retraining
- Data Science Capability
- IT Infrastructure

For an overview of the group's work and our 2022 programme, please read the [presentation](#) from our programme launch meeting on 9 February.

As a member you will be invited to our monthly meetings for expert presentations and discussions. You will also receive access to members websites and library, our members discussion forum, and will receive regular updates from our Theme Groups. All members are encouraged to contribute input, eg by sharing experience from their work through discussions and presentations.

Members are also invited to join one of our **Theme Groups**, small groups running knowledge exchange and research collaboration activities on key areas of our ML

### ML Group 2022 Monthly Meeting Presentations

Date	Speaker	Presentation
October 26	Florian Dumpert (Federal Statistical Office of Germany)	Workshop on Quality Aspects of Machine Learning - <a href="#">presentation slides</a>
	Javier Oyarzun, Laura Wile (Statistics Canada)	Quality Control of Machine Learning Coding: A Statistics Canada experience - <a href="#">presentation slides</a>
September 21	Yuhua Li (Cardiff University, UK)	Covariate shift detection based on exponentially weighted moving average ( <a href="#">presentation slides</a> )
	Riitta Piela (Statistics Finland)	Reaching for MLOps Level 1 at Statistics Finland ( <a href="#">presentation slides</a> )
August 31	Summer Wang (Australian Bureau of Statistics)	Raising Survey Response Rates by Using Machine Learning to Predict Gold Providers ( <a href="#">presentation slides</a> )
	Saeid Molladavoudi (Statistics Canada)	Statistics Canada's Framework for Responsible ML ( <a href="#">presentation slides</a> )
June 15	Piet Daas, CBS Netherlands	Using web site texts to identify different types of companies ( <a href="#">presentation slides</a> )
	David Corney, Full Fact, UK	How to stop people misusing statistics: Automatic verification of statistical claims ( <a href="#">presentation slides</a> )
May 4	Florian Dumpert (Federal Statistical Office of Germany)	Quality Framework for Statistical Algorithms ( <a href="#">presentation slides</a> )
April 6	Abel Dasyiva (Statistics Canada)	Estimating linkage errors without training data and without assumptions about the interactions among the linkage variables ( <a href="#">presentation slides</a> )
	Joep Burger (CBS Netherlands)	Convolutional neural networks for learning target variables and extracting image features from Earth ( <a href="#">presentation slides</a> )

- Coffee and Coding recordings
- Use cases from official statistics
- Information on ML Group 2022 activities & membership

[statswiki.unece.org/display/ML/Machine+Learning+Group+2022](https://statswiki.unece.org/display/ML/Machine+Learning+Group+2022)

# ML Group 2022 Webinar – 30 November

- Session I 1000-1130 CET
  - Applications of machine learning: web scraping data, text classification, imagery data, AIS data
- Session II 1500-1630 CET
  - Statistical production issues: quality of training data, model retraining, IT infrastructure
- Registration open on Eventbrite
  - Go to the link on the Machine Learning 2022 page [here](#)

# Thank you for joining us!

Email [ML2022@ons.gov.uk](mailto:ML2022@ons.gov.uk)

Website:  
[statswiki.unece.org/display/ML/Machine+Learning+Group+2022](https://statswiki.unece.org/display/ML/Machine+Learning+Group+2022)