Course Syllabus and Manual
for

Data Wrangling and Visualization

Course CODE

Spring 2026

Instructor: Krishanu Karmakar
kkarmakar@jgu.edu.in

Course Information
Course Duration: 15 weeks
Credit Hours: 4
Meetings: TBA
Location: TBA

Pre-requisites: At least one course of Statistics and the ability to operate Windows based computers.
(No knowledge of any specific statistical software needed)

Instructor Information
Education: PhD in Public Policy, MA in Economics, B.Sc. (Hons) in Economics
Office: 25 West, FOB, 6th Floor
Office Hours: TBA
Homepage:
https://jgu.edu.in/jsgp/faculty/dr-krishanu-karmakar

LinkedIn

YouTube

Keywords

Data Wrangling; Data Analysis; Data Visualization; Stata; NSS Survey; NSS Data.

Course Description

Data analysis and visualization are bread and butter for a career in either Economics or Public Policy. Data lets us create numbers and visualizations that tell compelling stories about human life, economy, and society. However, data cannot usually be used directly for analysis and for creating meaningful visualization. Collected data must be cleaned, recoded, new variables created, and combined with other data to make it suitable for analysis. In short data must be wrangled properly. This course intends to train students in the most popular and useful methods of data wrangling using the software Stata. This course also intends to teach students how to create good visual

representation of data by training them in the theory and current practices of data visualization.

Expected Learning Outcomes

After successfully completing this course, the students are expected to be able to

1. Work with the software Stata.
2. Manage and wrangle datasets using the software Stata.
3. Handle messy data and convert that into tidy data in Stata.
4. Create certain types of meaningful and clean diagrams in Stata following the current best practices in Data Visualization.
5. Extract and manage NSS data using Stata.
6. Do rudimentary analysis of NSS data using Stata.

Students differ on the preparation, effort, and interest. Therefore, student to student variation in achievement of these expected learning outcomes is natural and expected.

Prerequisite

Students will be expected to understand basic statistics and comfortable with using computers. Students also need access to personal laptops. ***Any prior knowledge of Stata is not needed***. (OP JGU Library has subscription of Stata and will facilitate installing the software in your laptops).

Teaching Activities:

1. Class lectures
2. Practical Demonstration
3. In Class Practice

Assessment

1. **Internal Assessment (60%):**
   a. **Best 2 out of 3 in class Tests** (each out of 30 marks)
2. **External Assessment (40%): One in-class examination** held during the examinations week. (1.5 hr, 40 marks)

Recommended/Required Book

There is no single textbook for this course. However, the following books will be useful in following the course material (To be completed)

Attendance, Missed, and Late Assignment Policy

University attendance policy will be adhered to in this course. No request of additional relaxation can be entertained during or at the end of the semester.
Alternative assessment will be provided as per the University regulations. The form of the alternative assessment will be solely at the discretion of the faculty.
Missed Assignment: Most of the time I shall decide on the merit of the case. I shall be considerate to things which are beyond your control. But empty excuses will not be condoned. Late Assignments will be considered missed assignments unless submitted with my prior approval.

Academic Honesty & Class Decorum

Cases of academic dishonesty including **plagiarism** (which is defined as more than 10% of your submissions being someone else's work, attributed or unattributed) will be handled according to university policy.

You will be asked to leave the class for any type of **disruptive behaviour**.

Repeated entry and exit (without a serious medical condition or with the written recommendation from DSC/Sukoon) will lead to you being asked to leave the class.

Usage of mobile phones is not allowed inside the class. If you use mobile phones in class, you will be asked to leave the class. **If you are expecting to receive an important call then do not come to class.**

Use of Generative AI except in Final Examination

You may intelligently use generative AI when applicable and possible. But you are required to indicate where you have used Generative AI clearly. One way of doing it is using footnote system in your write up which includes the declaration of usage of generative AI, the engine that you used, and the list of prompts that you used. However, mindless application of generative AI (for example, but not limited to, verbatim copy of AI generated material) and not disclosing the usage of generative AI will be punished as low-quality work.

For final examination University regulation will supersede the above scheme.

Special Needs

If you have any special needs that I should be aware of and should be considered in relation to your examinations and assignment submissions, you should talk to me and get yourself registered with the disability committee. The disability committee need to inform me about what type of accommodation you should be provided with.

Tentative Course Outline & Session Plans

Day	Topics
Aug	
1	**Introduction to data management from the perspective of replicability** **&** **Tidy data**
2	**Introduction to Stata:** • GUI, • Working folder, • Modes of operation, • Menus, • do file
3	**Introduction to Stata (contd.):** • loading .dta & non-.dta data • basic exploration • basic descriptive statistics • basic frequency distributions
4	**Introduction to Stata (contd.):** • loading non-.dta data • help system • Command syntax of Stata
5	**Data wrangling 1:** • creating variables (generate-replace & egen) • renaming • recoding
6	**Data wrangling 2:** • labeling • notes • working with string variables • tostring-destring
7	**Test 1**
8	**Fixed Format Data:** • structure, • dictionary file,

Day	Topics
	• extraction
Sept	
9	**NSSO data:** • structure • dictionary file
10	**NSSO data:** • dictionary file • extraction
11	**Data management:** • local & global macro • foreach & forvalues loop
12	**Data management:** • foreach & forvalues loop
13	**Wrangling NSSO Data in Stata:** • Appending • Merging • Reshaping
14	**Wrangling and Analysis of NSSO Data in Stata 1:** • Creating new variables (including weights and recodes) • Labels and notes • Basic analysis
15	**Wrangling and Analysis of NSSO Data in Stata 2:** • Basic Analysis (emphasize the circular process of wrangling and analysis) • Creating new variables (including weights and recodes) • Labels and notes • More analysis
16	**Test 2**
17	**Basic Graphics in Stata:** • Bar chart • Histogram • Box & Whisker Plot
Oct	
18	**Basic Graphics in Stata:** • Scatterplot • Line plot
19	**Principles of Data Visualization:** • Visual perception & Pre-attentive processing • Gestalt Principles
20	**Principles of Data Visualization:** • Elements of graphs • Encoding data in graphs

Day	Topics
	• Encoding relationships
21	**Principles of Data Visualization:** Wrap up
22	**Advanced Graphics in Stata: Applying Principles of Data Visualization**
23	**Advanced Graphics in Stata: Applying Principles of Data Visualization**
24	**Advanced Graphics in Stata: Applying Principles of Data Visualization**
25	**Test 3**
Nov	
26	Working with String Variables in Stata
27	Powerpoint

This is a tentative schedule; topics might change a little for pedagogical reasons.