**Data Quality Assurance and Quality Control**

**EFB 202 – Ecological Monitoring and Biodiversity Assessment**

**Learning Objectives:**

At the end of the day, students will be able to:

1. Actively contribute to data collection via development of quality assurance protocols
2. Actively contribute to data analysis via data entry and quality control
3. Organize and display GPS locations using Garmin BaseCamp and ArcGIS Pro

**Outline:**

1. Context
   1. Why am I studying Data QA/QC?
2. Introduction
   1. What is Data QA/QC?
3. Application
   1. How do I apply Data QA?
      1. Exercise 1: Soil–Plant Relationships
      2. Exercise 2: Crayfish Habitats
   2. How do I apply Data QC?
      1. Exercise 3: Deer Herbivory
      2. Exercise 4: GPS Coordinates
4. Storage
   1. Where should I keep my QA/QC Data?
5. Access
   1. When should I share my QA/QC Data?
   2. Who provides QA/QC Data that I can use?
6. GPS Instructions

**I. Context**

1. Primary goals of this class are field methods and species identification
2. Other modules taught to apply these skills are Science Inquiry, Sampling, and Statistics
3. Data QA/QC bridges the gap between science inquiry/sampling and statistics

**II. Introduction**

1. What is data QA/QC?
2. Quality assurance (QA) is used to prevent defects
3. Quality control (QC) is used to identify defects
4. Many regulatory agencies require data management plans (DMPs)

**IIIa. Application of QA**

1. Handwritten vs. digital data collection
2. Developing a data collection plan

**Exercise 1: Soil–Plant Relationships**

**A person in a forest

Description automatically generated**Your EFB 202 research team decides to examine the relationship between soil characteristics and plant community composition. Specifically, you intend to examine the relationship between moisture and pH and plant biodiversity at different sites around Cranberry Lake Biological Station. To control for potential differences across CLBS, your team designs a symmetrical grid of 20 points. Your plan is to assess plant diversity at each point and collect soil samples for lab analysis. What information do you need to collect while you’re in the field?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Exercise 2: Crayfish Habitats**

Your EFB 202 research team decides to examine the relationship between aquatic habitat characteristics and the abundance of crayfish. Your team selects 27 locations along the shoreline of Cranberry Lake Biological Station and Barber Island. Each night, you will deploy crayfish traps baited with anchovies and count the number of crayfish captured. You will also assess the aquatic habitat characteristics at each site. What information do you need to collect while you’re in the field?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A group of lobsters in muddy water

Description automatically generated

**IIIb. Application of QC**

1. Principals of Data Entry – Formatting your Excel sheet for analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

1. Metadata sheets / README files – guidelines for making data accessible and usable

**Exercise 3: Deer Herbivory – Part 1**

For the past few years, students have been collecting data on deer herbivory and plant diversity during the botany module of EFB 202. These students collect the data at the same predetermined and numbered sites each year. You will be given datasheets from the study and asked to enter the data into Excel for inclusion in the larger dataset.

Once you’ve input all your data into Excel, copy and paste your data into the Google Sheet that was shared with you so that we all have access to the complete dataset.

**Exercise 3: Deer Herbivory – Part 2**

Now that you have the complete dataset, discuss with your research team what you can do to identify errors. Are there any differences between your team’s dataset and the other dataset?

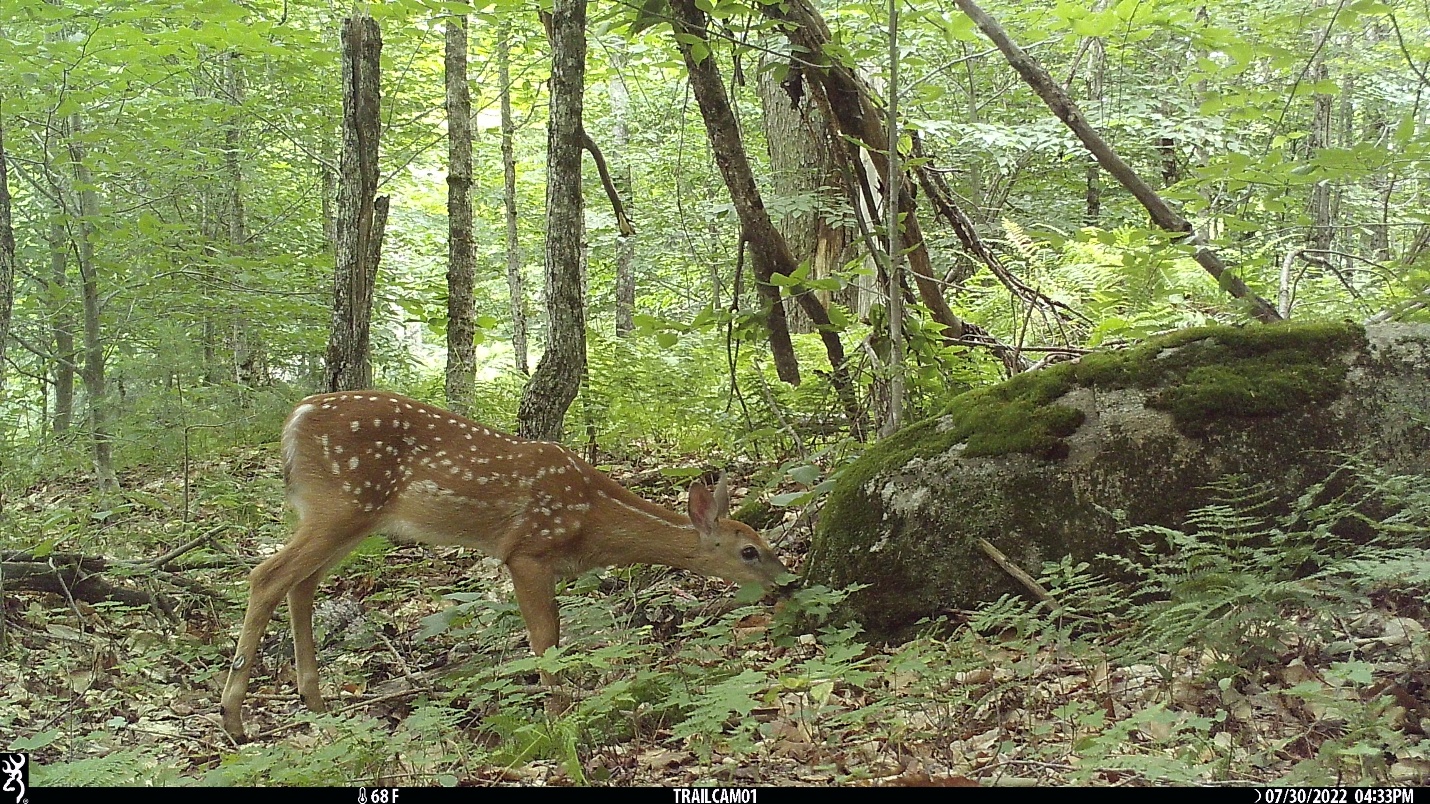
**A group of people in the woods

Description automatically generated**

1. Common Sources of Error
2. Cleaning Your Data:
   1. Is the total number of columns correct?
   2. Is the total number of rows correct?
   3. Are there any abnormal values?
   4. Are there any cells in the dataset that are blank?
   5. What basic statistics can be used to identify errors?
3. Plotting GPS Coordinates

**Exercise 4: GPS Coordinates**

In 2022, a camera trap array was established at Cranberry Lake Biological Station. The coordinates for the grid centers and the camera traps were saved to the GPS units you’ve been given. The goal of this exercise is to create a GIS map of those coordinates to make sure they’re in the right places and to share as part of a presentation. To complete this exercise, follow the instructions at the end of this handout in Part VI.



1. What to do when the data is wrong but you can’t fix it?

**IV. Storage**

1. How to name files for storage
2. Multiple copies and external hard drives
3. Cloud-based storage

**V. Access**

1. Who has access to the data you collected?
2. When is it appropriate to share access to your data?
3. Where can I find publicly available data?
   1. GBIF (global biodiversity information facility)
   2. IUCN Red List
   3. US Federal Government
   4. Movebank
4. What about data from specific projects?
5. Where is project data often stored?
   1. Scientific Journals
   2. Research Gate
   3. GitHub
   4. Digital Commons
6. How to interpret public data
   1. README files
   2. Applying Data QC

**VI. GIS Instructions**

To create your ArcGIS map from locations saved on a GPS, follow these steps:

1. Plug the GPS unit into your computer using the USB transfer cable
2. Open the Garmin BaseCamp application
3. Select “Device” -> “Receive from Device” -> choose your device from the menu -> select “OK”
4. Your locations should automatically appear on the map; if they don’t, you can Zoom to your points by selecting the folder you imported titled “Data received from…”, right-clicking on the folder, and selecting “Show on Map”, or you can manually zoom to CLBS
5. To clean up your locations, select “Garmin\_Asia”, “Garmin\_Europe”, “Garmin\_USA” and any other layers you don’t want exported to ArcGIS. You can select multiple layers at the same time by holding “ctrl” on PCs or “command” on Macs. Then, right click on one of those layers and select “Delete”
6. To export the locations to ArcGIS, highlight the folder you want to export, then select “File” -> “Export” -> the file name
7. Save the file using the name and location of your choice as a .gpx file
8. Open a new map in ArcGIS
9. Select “Tools” on the Analysis tab to open the Geoprocessing window
10. Search for the “GPX to Features” tool
11. Select your Input GPX File from where you saved it and rename the Output Feature Class whatever you find most useful. The Output Type should be “Points”; select “Run”
12. Your points should appear on the ArcGIS Map. If you can’t see them, you can right click the layer in the left side Table of Contents and select “Zoom to Layer”
13. Next, search for the “Add XY Coordinates” tool in the Geoprocessing window
14. Input the point feature that you just created; select “Run”
15. Right click your feature in the Table of Contents, then select "Attribute Table”; the last three columns of the attribute table should be “POINT\_X”, “POINT\_Y”, and “POINT\_Z”, which are the latitude, longitude, and elevation of each point, respectively
16. You can export these coordinates to Excel using the “Table to Excel” geoprocessing tool, which allows you to add them to your dataset without editing them or retyping them manually
17. You can manually modify the points’ aesthetics by clicking the point symbol underneath your feature layer in the Table of Contents
18. You can manually alter the map aesthetics by selecting “Basemap” on the Map tab and choosing your preferred map