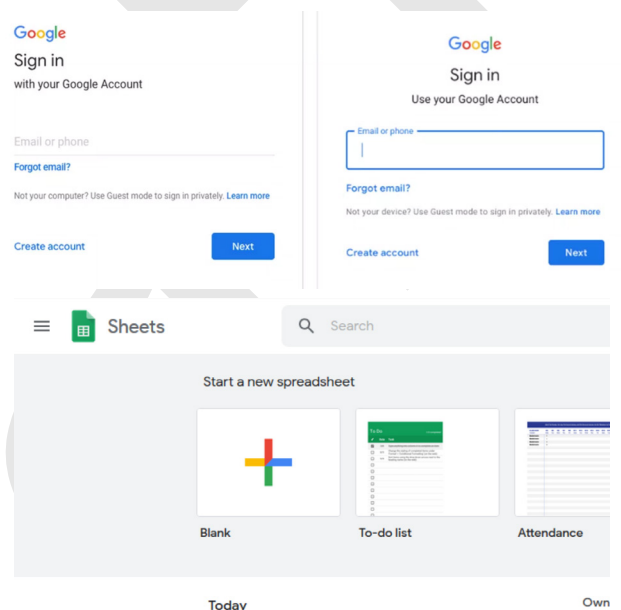## Google Sheets: ImportHTML as Tables

Welcome to the Essential Data Science Data Cleaning Series. This series supplements the Iowa State University Geospatial Technology Training Program and is designed to provide quick, easy instructions for creating interactive data visualizations. Google Data Studios is a free web based application that incorporates data analysis and visualization.

In this tasksheet you will learn how import HTML tables in Google Sheets, a free web based spreadsheet program similar to Microsoft Excel. Using a simple formula you will create a table in Google Sheets with the URL of a website. Google Sheets will automatically format the table data, leaving you with very little work left to do.

1. Getting Started with Google Sheets

a.   First you will need to sign in or create a free Google account.

b.   You can access Google Sheets here **https://www.google.com/sheets/about/** once you are signed into your account.

c.   Click on **Blank** to start a new spreadsheet.

2. Creating Import Formula

a.   Open the following link to observe the data used in this tasksheet **https://en.wikipedia.org/wiki/List_of_counties_in_Iowa**.

b.   You will notice a long table listing Counties in Iowa, along with some information about each county.

c.   **Copy** the **URL** for the Wikipedia page **https://en.wikipedia.org/wiki/List_of_counties_in_Iowa**.

d.   The formula for data wrangling is =IMPORTHTML("url", "query", "index").

•   URL - Is the website link you will copy.

•   Query - can be either List or Table, it tells Google what to build.

•   Index - Tells Google what Table or List to import from that webstie.

e.   In Goole Sheets click on cell **A-1** on your blank spreadsheet and enter the formula **=IMPORTHTML("https://en.wikipedia.org/wiki/List_of_counties_in_Iowa", "table", "2")** and press **Enter**.
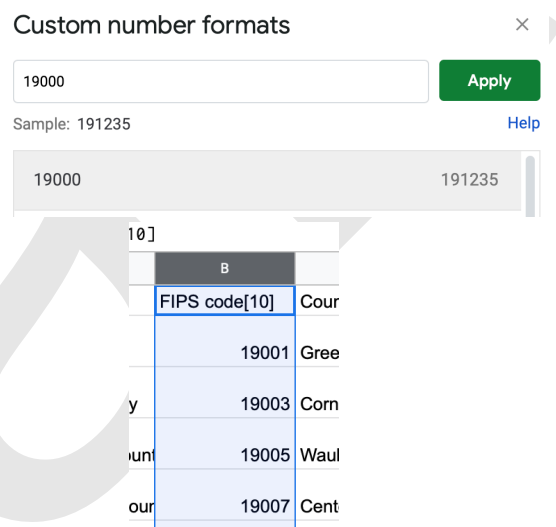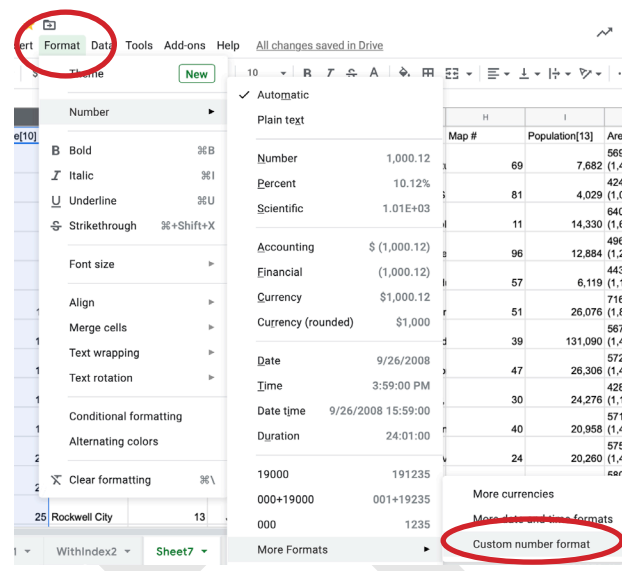
f. The table of Counties will fill out the spreadsheet. This is the second table on the Wikipedia page, and is why you used an index of 2.
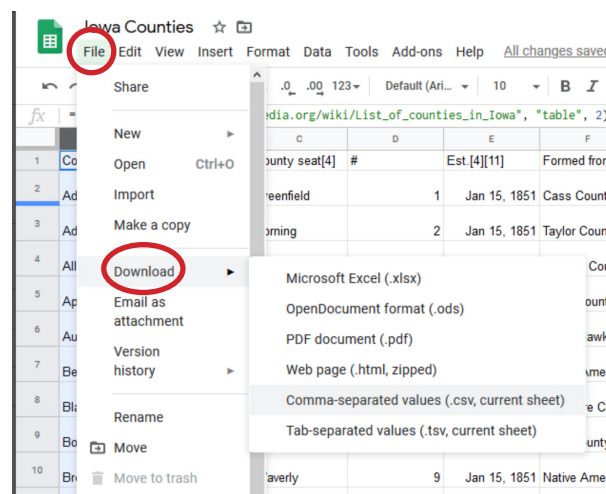
3. Custom Table Formating

a. You will notice your FIPS codes have been reformatted and no longer have additional zero's. Fix this by selecing **Columb B** so all its cells are highlighted.

b. From the main menu open **Format** then expand **Number** then expand **More Formats** and select **Custom Number Format**.

c. The **Custom Number Formats** window opens, here you can select commonly used number formats or type in your own. Since Iowas FIPS code is 19 type **1900** in the Custom Number Format box and click **Apply**.

d. Google Sheets will automatically number the counties with their proper FIPS Code.

e. You can also experiment with different index numbers. Click cell A-1 and replace your index value of 2 with **1** and click **Enter**.

f. Your table will change to display the abbrevated list of counties found at the top of the wikipedia page.

4. Downloading Tables

a. Once you are done editing your imported table you can download the table to your computer to be used with other software.

b. Give your spreadsheet a title by clicking **Untitled spreadsheet** at the top and typing **Iowa Counties**, and press **Enter**

c. Open **File** from the main menu and expand **Download**. You have a few options of file type to download as such as: CSV, Excel xlsx, PDF, and more.

d. You work is automatically saved by Google Sheets, so you may simply close the program when finished.

Contact:

Professor Christopher J. Seeger, ASLA, GISP cjseeger@iastate.edu, 515-509-0651 or Bailey Hanson bahanson@iastate.edu, 515-520-1436 for more information about the Geospatial Technology Program. This tasksheet and more are available at **www.extension.iastate.edu/communities/gis**