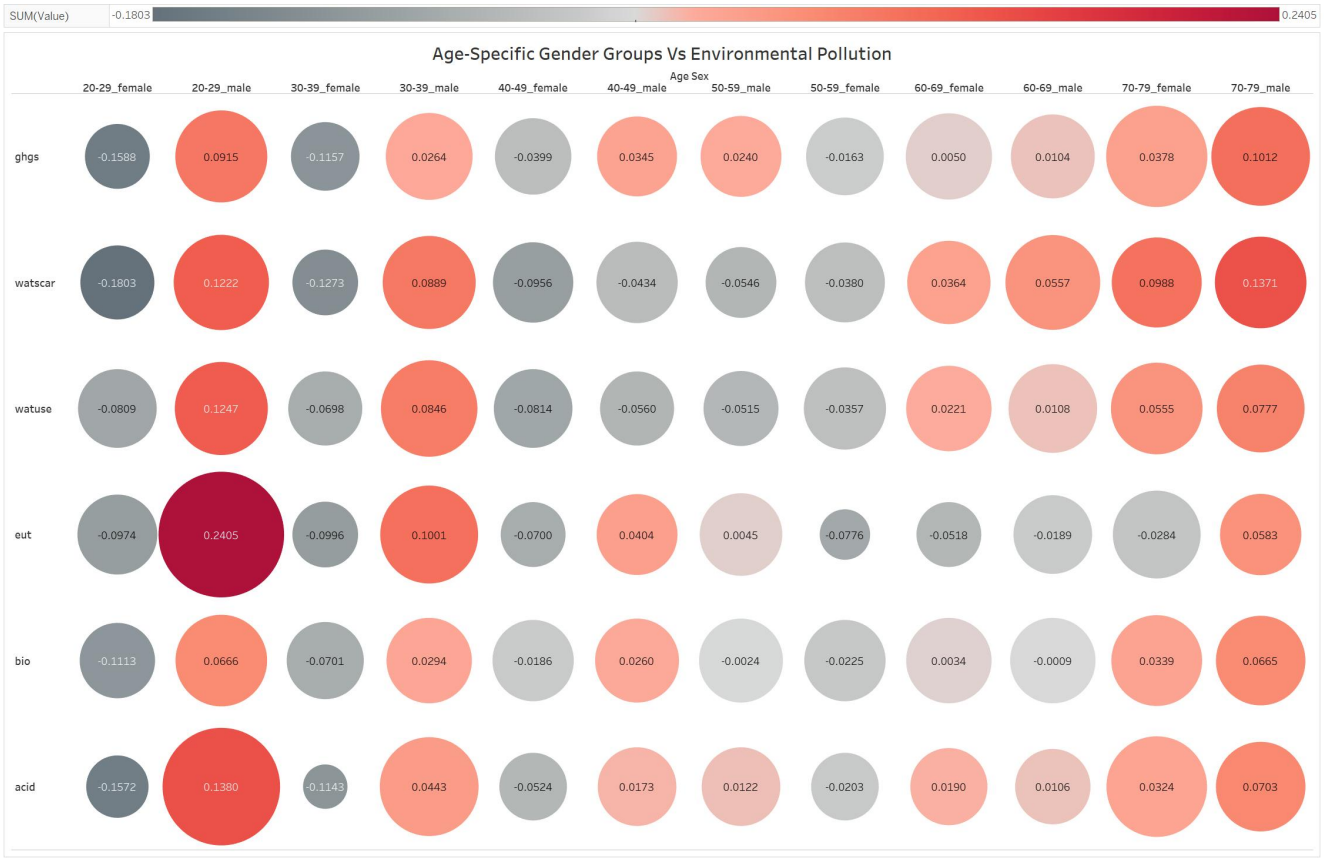
- **Image:** Age-Specific Gender Groups Vs Environmental Pollution



- **Visual Design Type:** Matrix Bubble Chart
- **Name of Tool:** Tableau
- **Age-Sex Groups**: A total of 12 groups, combining different age ranges and genders.
- **Variables:**
    1) ghgs: GHG emissions (GreenHouse Gas) measured in kg(mean & standard deviation)
        a. The reason for selecting GHG emissions is that this characteristic includes all greenhouse gases, eliminating the need to integrate the values of N2O and CH4.
    2) watscar: Water Scarcity(mean & standard deviation)
        a. The feature of Water Scarcity was selected because it reflects the impact on water resources.
    3) watuse: Agricultural Water Usage in cubic meters(mean & standard deviation)
        a. Agricultural Water Usage was selected because it can also reflect the impact on water resources.
    4) eut: Eutrophication Potential - measured in g of PO4e, gPO4e(mean & standard deviation)
        a. Eutrophication Potential was selected because it is one of the environmental pollution assessment indicators and cannot be substituted by other metrics.
    5) bio: Biodiversity Impact - species extinction per day(mean & standard deviation)
        a. Biodiversity Impact was selected because it is one of the environmental pollution assessment indicators and cannot be substituted by other

metrics.

6) acid: Acidification Potential(mean & standard deviation)
    a. Acidification Potential was selected because it is one of the environmental pollution assessment indicators and cannot be substituted by other metrics.

- **Visual Mappings:**
    1) The **x-axis** represents different Age-Sex sample groups, while the **y-axis** represents various environmental pollution assessment indicators.
    2) The **colour** of the bubbles represents the pollution level of the current sample group for the specific environmental pollution assessment indicator (the redder it is, the more severe the pollution).
    3) The **numerical labels** on the bubbles present the pollution level of the current group in a more precise manner.
    4) The **size** of the bubbles indicates the standard deviation of the current pollution level data (the larger the bubble, the greater the standard deviation).
    5) The bubbles were chosen to have an aesthetically pleasing and harmonious **round shape**.
    6) The design of the **x-axis** incorporates a hierarchical structure, first grouping by age range and then by gender.
- **Unique Observation:**
    1) From the visualization we can see, for the same age range, male diets generally cause more environmental pollution than female diets.
    2) For female, as they age, the pollution caused by their diets tends to increase. For male, the pollution decreases until around age 50, after which it follows a similar upward trend as for female.
    3) The 20-29 male sample group has a higher standard deviation in its eutrophication (eut) and acidification (acid) data (as indicated by the noticeably larger bubble size). This suggests a higher level of data dispersion and a greater likelihood of outliers.
    4) Through interactive features such as dragging, the hierarchy can be adjusted to analyze different data points (as demonstrated in the Interactive Demo video).
- **Data Preparation:**
    1) Python, along with the pandas and sklearn libraries, was used for standardizing the necessary data, such as the mean and standard deviation, to properly display them on an information visualization chart.
    2) The age_group and sex features were combined into a single feature called age_sex.
    3) The wide-format data was converted to long-format data to facilitate creating the appropriate chart using Tableau.
- **URL** to **Interactive Demo:**
    https://youtu.be/thlKGfosZJM
- **URL to source code:**
    Python code: https://github.com/AndrewMaaay/RM-CW2.git