Slugs and Snail Data Analysis and Visualization

Questions:

1) How do the number of slugs and snails found by biologists change over time in this data set?
2) How do the number of slug and snail species found by biologists change over time in this data set?

Pretty plot to answer the questions:

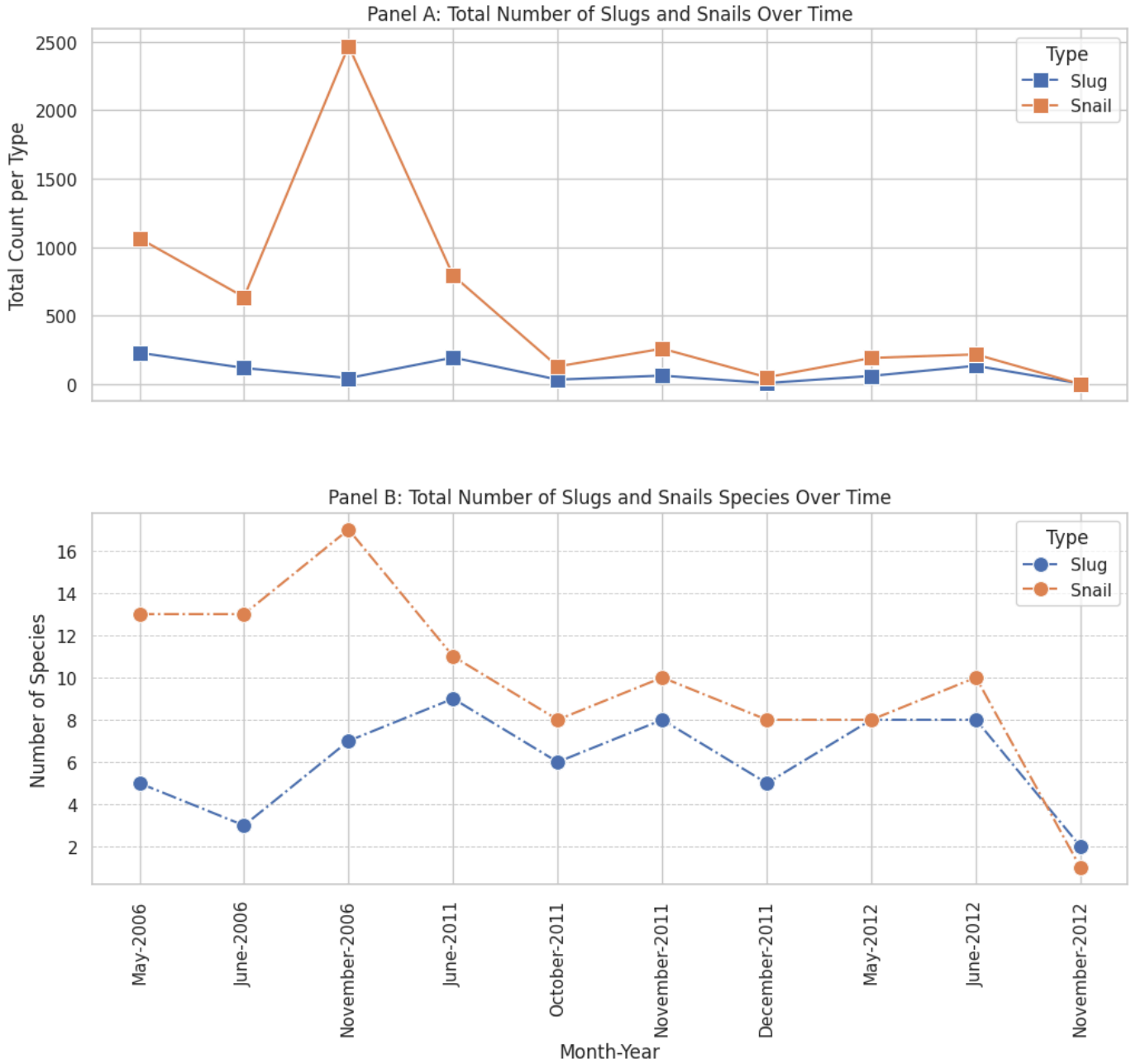

Figure 1. Two lineplots showing the changes in slug and snail counts between May 2006 and November 2012. Panel A shows the total counts of slugs and snails, with snail counts peaking dramatically in November 2006, while slug counts remain relatively stable. Panel B illustrates the number of slugs and snails species, highlighting the high variability and significant peak for snail species in November 2006, contrasted with consistently lower and more stable slug species counts.

Illusion plot for answering the questions:

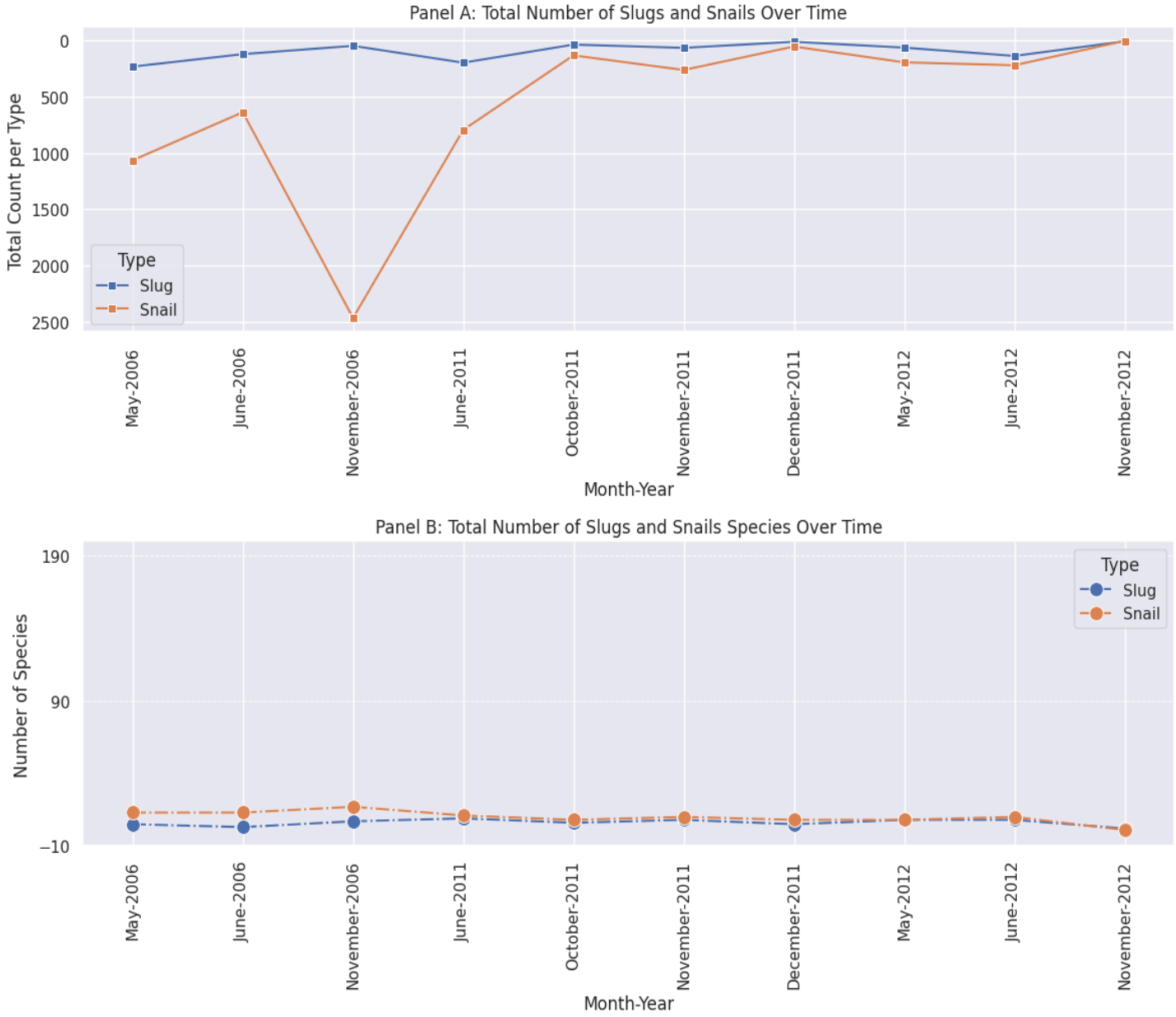

Figure 2. Two lineplots showing the changes in slug and snail counts between May 2006 and November 2012.Panel A shows the total counts of slugs and snails, where snail counts vary more significantly, particularly with a noticeable dip in November 2006.Panel B illustrates an overall similar number of slugs and snails species.Despite the seemingly different trends in total counts by type(Panel A) and number of species per type(Panel B), there is an illusion caused by the data representations(inverted y-axis in Panel A) and different scales(Panel B).

Code snippet and steps to answer the questions:

In order to answer the 2 questions, I had to perform data preprocessing steps like data cleaning, data manipulation and finally perform data visualization to understand any pattern or trend.

Step 1: Subset the data by columns to choose only those columns which will help in answering the dataset associated questions.
Step 2: Perform data cleaning by changing the datatype of "Date" from "Object" to "DateTime".
Step 3: Adding a new column "Year" by extracting the "year" value and another column "Month" by extracting "month" from "Date".
Step 4: Adding the "Season" column based on the month of the year.
Step 5: Drop the "Date" column as it does not contribute for further analysis.
Step 6: Perform appropriate grouping and aggregation pandas operations to answer the two questions.
Step 7: Finally, develop an appropriate plot for data visualization.

Code snippet for **Pretty plot to answer question 1 and 2:**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#Creating a dataframe that contains only lugs and snails found by the
biologists.
slugs_snail_df = filtered_df[(filtered_df['Type']=='Slug') |
(filtered_df['Type']=='Snail')]
number_slugs_snail = slugs_snail_df.groupby(['Year',
'Month','Type'])['Count'].sum().reset_index(name='Total')

different_species_slugs_snail = slugs_snail_df.groupby(['Year', 'Month',
'Type'])['ScientificName'].nunique().reset_index(name='TotalSpecies')

#Pretty Plot
months_order = ['January', 'February', 'March', 'April', 'May', 'June',
'July', 'August', 'September', 'October', 'November', 'December']

# Define the correct order of months for number_slugs_snail
number_slugs_snail['Month'] = pd.Categorical(number_slugs_snail['Month'],
categories=months_order, ordered=True)
```

```
# Sort by year and then by month
number_slugs_snail = number_slugs_snail.sort_values(by=['Year', 'Month'])

# Create a new column for plotting purposes
number_slugs_snail['Month-Year'] = number_slugs_snail['Month'].astype(str)
+ '-' + number_slugs_snail['Year'].astype(str)
# Define the correct order of months for species_slugs_snail
different_species_slugs_snail['Month'] =
pd.Categorical(different_species_slugs_snail['Month'],
categories=months_order, ordered=True)
# Sort by year and then by month
different_species_slugs_snail =
different_species_slugs_snail.sort_values(by=['Year', 'Month'])

# Create a new column for plotting purposes
different_species_slugs_snail['Month-Year'] =
different_species_slugs_snail['Month'].astype(str) + '-' +
different_species_slugs_snail['Year'].astype(str)

sns.set(style="whitegrid")
# Create subplots with division bars
fig, axs = plt.subplots(2, 1, figsize=(12, 12*0.618), sharex=True,
gridspec_kw={'hspace': 0.3}, frameon=True)

# First subplot
sns.lineplot(ax=axs[0], data=number_slugs_snail, x='Month-Year',
y='Total', hue='Type', marker='s', markersize=10)

# Customize the first subplot
axs[0].set_title('Panel A: Total Number of Slugs and Snails Over Time')
axs[0].set_ylabel('Total Count per Type')
axs[0].legend(title='Type')
axs[0].tick_params(axis='x', rotation=90)
axs[0].grid(True)

# Second subplot
sns.lineplot(ax=axs[1], data=different_species_slugs_snail,
x='Month-Year', y='TotalSpecies', hue='Type',
marker='o',markersize=10,linestyle='-.')
```

```
# Customize the second subplot
axs[1].set_title('Panel B: Total Number of Slugs and Snails Species Over
Time')
axs[1].set_xlabel('Month-Year')
axs[1].set_ylabel('Number of Species')
axs[1].legend(title='Type')
axs[1].tick_params(axis='x', rotation=90)
axs[1].grid(True, axis='y', linestyle='--', linewidth=0.7)

plt.tight_layout()
plt.show()
```

Code snippet for **Illusion plot to answer question 1 and 2:**

```
# Illusion Plot
sns.set(style="darkgrid")

# Create subplots with division bars
fig, axs = plt.subplots(2, 1, figsize=(12, 12*0.618))

# First subplot
sns.lineplot(ax=axs[0], data=number_slugs_snail, x='Month-Year',
y='Total', hue='Type', marker='s')
# Add distortion to y-axis
axs[0].invert_yaxis()
# Customize the first subplot
axs[0].set_title('Panel A: Total Number of Slugs and Snails Over Time')
axs[0].set_ylabel('Total Count per Type')
axs[0].legend(title='Type')
axs[0].tick_params(axis='x', rotation=90)
axs[0].grid(True)

# Second subplot
sns.lineplot(ax=axs[1], data=different_species_slugs_snail,
x='Month-Year', y='TotalSpecies', hue='Type',
marker='o',markersize=10,linestyle='-.')

# Customize the second subplot
axs[1].set_title('Panel B: Total Number of Slugs and Snails Species Over
Time')
axs[1].set_xlabel('Month-Year')
```

```
axs[1].set_ylabel('Number of Species')
axs[1].legend(title='Type')
axs[1].tick_params(axis='x', rotation=90)
axs[1].grid(True, axis='y', linestyle='--', linewidth=0.7)

# Set y-axis scaling from -10 to 200 with a gap of 100 each for the second subplot
axs[1].set_ylim(-10, 200)
axs[1].set_yticks(np.arange(-10, 200, 100))

plt.tight_layout()
plt.show()
```

Illusions and potential issues with the plot.

- **Inverted Y-Axis Scale:** In Panel A, the y-axis does not start at zero as it is inverted, which can give a misleading impression of the magnitude of changes over time between Slug and Snail counts. This makes it difficult to interpret the true counts and trends accurately.
- **Overlapping Lines:** In Panel B, the overlapping lines for Slug and Snail make it hard to differentiate between the two types and their trends. This scaling issue can obscure any real differences or patterns in the data.
- **Negative Values in Count Data:** In Panel B , y-axis has negative values which is not logical for count data (as counts cannot be negative). This might suggest an issue with data processing or plotting.

To improve the plots, consider:

- Ensuring the y-axis starts at zero to provide a clearer comparison.
- Avoiding negative values on the y-axis for count data.
- Use appropriate scaling to get an informative trend for each species of the snail and slug type.