YouTube Data Analysis — Debugging Summary (2025-06-19)

This document summarizes all **debugging issues, root causes, and fixes** from today's session on your YouTube category analysis project.

---

1 `trend.sum(axis=1)` — TypeError

Error Message

```
TypeError: unsupported operand type(s) for +: 'float' and 'str'
```

Cause

- The `trend` DataFrame included non-numeric columns (e.g., strings or datetime), which can't be summed row-wise.

Solution

Filter for numeric columns before summing: python `trend_numeric = trend.select_dtypes(include='number')` `total_watched_times = trend_numeric.sum(axis=1)`

---

2 `'Series' object has no attribute 'select_dtype'`

Error Message

```
AttributeError: 'Series' object has no attribute 'select_dtype'
```

Cause

- `.select_dtypes()` only works on DataFrames, not Series.

Solution

Use it on a DataFrame like: python `category_numeric = category_df.select_dtypes(include='number')`

---

3 'count' column not found

Error Message

```
KeyError: 'count'
```

Cause

- `groupby(...).size()` returns a Series, not a DataFrame with a named column.

Solution

Convert to DataFrame and name the column: python df_counted = df.groupby(['year_month', 'category']).size().reset_index(name='count')

---

4 No axis named 1 for object type Series

Error Message

```
No axis named 1 for object type Series
```

Cause

- This happens when `divide(..., axis=1)` is mistakenly applied to a Series instead of a DataFrame.

Solution

Ensure axis=0 is used: python percent_df = category_numeric.divide(total_watched_times, axis=0) * 100

---

5 KeyError: '1   ' in color mapping

Error Message

```
KeyError: '1            '
```

Cause

- `custom_colors` used category names as keys, but `percent_df.columns` contained channel names.

Solutions

- Make sure keys in `custom_colors` match the column names:
  - If `percent_df` uses **channel names**, define `custom_colors` with channel keys.
  - If using **categories**, build `percent_df` from a category-based pivot table.

---

6. '   ' category dominates the chart

Issue

- '   ' often contains many unmatched channels and overwhelms other categories.

Solutions

- Option 1: Drop the column entirely: python `percent_df = percent_df.drop(columns='          ')`

- Option 2: Review and reassign unknown channels: python `df[df['category'] == '          ']['channel'].value_counts()`

References

- [Pandas: groupby documentation](#)
- [Pandas: select_dtypes](#)
- [Matplotlib: Horizontal Bar Chart](#)