```python
In [1]: import sys
        from pathlib import Path

        # Add project root to Python path
        ROOT = Path.cwd().parent   # adjusts if notebook is in /notebooks/
        sys.path.append(str(ROOT))
```

```python
In [2]: from pathlib import Path
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        sns.set(style="whitegrid", context="notebook")

        from src.cleaning import make_processed

        df = make_processed()  # runs pipeline + returns cleaned df
        df.head()
```

Saved cleaned data to /Users/sherman/Library/Mobile Documents/com~apple~CloudDocs/SUTDie/Term 7/02.218TS Intro to Psychology/IntroPsych/data/proce
ssed/study_results_clean.csv

Out[2]:

	participant	group	correct_1	correct_2	correct_3	correct_4	correct_5	correct_6	correct_7	correct_8	...	abs_19	cws_19	p_20	abs_20	cws_20
0	CG1	CG	0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.000000	0.400000	1.000000	0.000000	0.000000
1	CG2	CG	0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	...	0.500000	0.600000	0.000000	1.000000	0.400000
2	CG3	CG	0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	...	1.500000	1.000000	1.000000	0.000000	0.000000
3	CG4	CG	1	1.0	0.0	1.0	1.0	1.0	1.0	1.0	...	1.500000	1.000000	0.833333	0.305556	0.066667
4	CG5	CG	1	1.0	0.0	0.0	0.0	1.0	1.0	1.0	...	0.555556	0.133333	0.500000	0.750000	0.200000

5 rows × 107 columns

```python
In [3]: df.describe()[["accuracy", "mean_conf", "total_abs", "total_cws"]]
```

Out[3]:

	accuracy	mean_conf	total_abs	total_cws
count	42.000000	42.000000	42.000000	42.000000
mean	0.523810	4.501190	21.207011	11.807937
std	0.143237	1.127726	2.152918	1.920942
min	0.200000	1.000000	15.694444	8.333333
25%	0.450000	3.925000	19.854167	10.666667
50%	0.525000	4.450000	21.319444	11.966667
75%	0.600000	5.225000	22.479167	12.800000
max	0.850000	6.400000	25.777778	16.533333

```python
In [4]: for col, title in [
            ("accuracy", "Distribution of accuracy"),
            ("mean_conf", "Distribution of mean confidence"),
            ("total_abs", "Distribution of total ABS"),
            ("total_cws", "Distribution of total CWS"),
        ]:
            df[col].hist()
            plt.title(title)
            plt.xlabel(col)
            plt.ylabel("Count")
            plt.show()
```