

CODE ARCHITECTURE — SYNTHDATA-REVIEW

Dependency map: code/

17 scripts, three pipelines, one shared core. `majority_vote.py` turns three model annotation workbooks into `majority_vote_annotations.xlsx`, which everything downstream reads. From there the graph splits into an agreement/kappa pipeline, a temporal-trend pipeline, and a facet-distribution (pie) pipeline — the last two feeding the summary figures at the right.

NODE FILL

raw / generated data file

shared core (`majority_vote.py`)

kappa / agreement pipeline

temporal-trend pipeline

facet / pie pipeline

NODE SHAPE

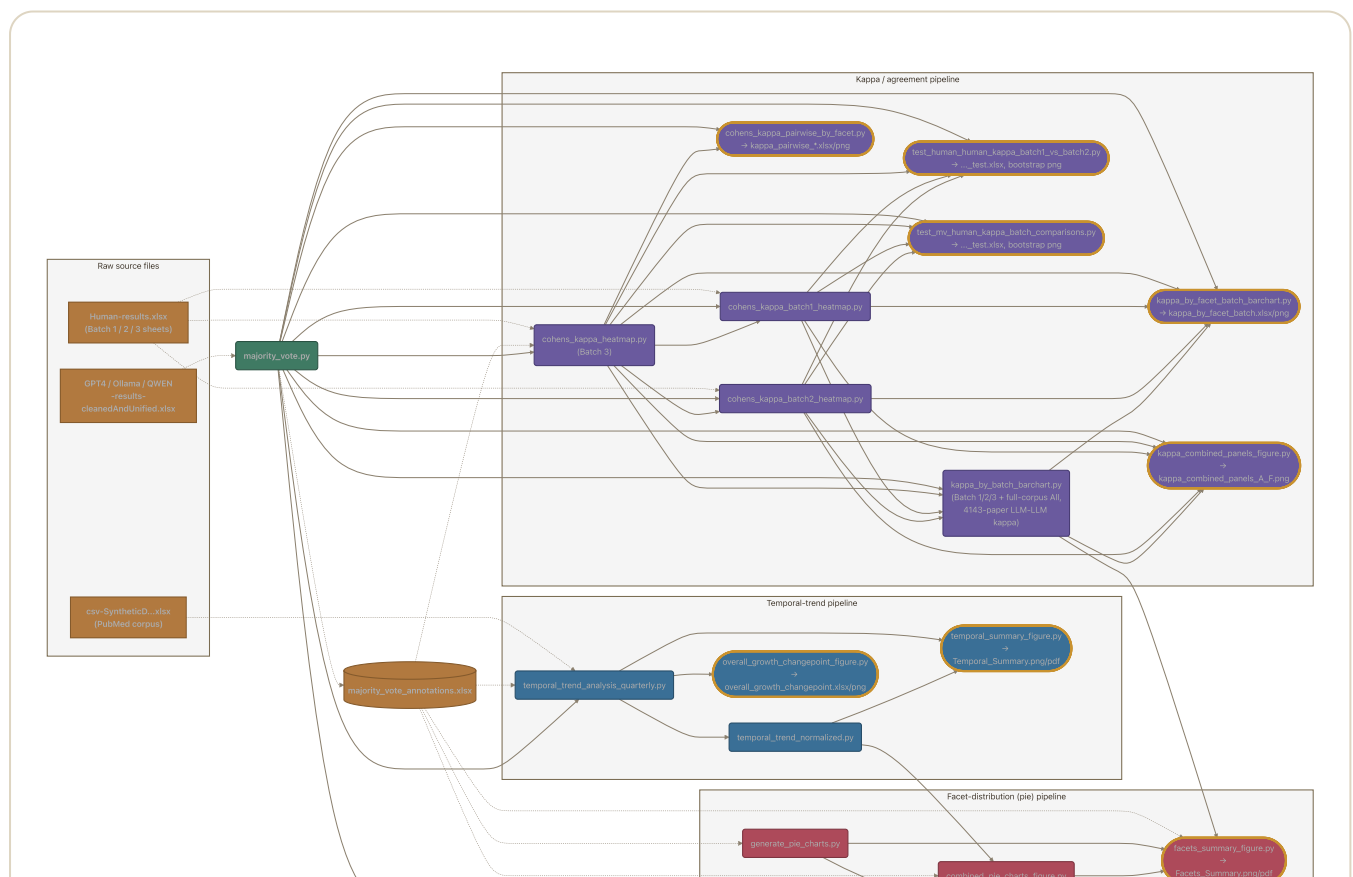
rounded — module, imported by others

 gold ring — terminal script (run it directly for output)

EDGE

 solid — Python import

 dashed — reads / writes a data file



Full reference – shortest script first

SCRIPT	LINES	READS
temporal_summary_figure.py	133	temporal_trend_candidates_qu temporal_trend_normalized.xl
facets_summary_figure.py	138	majority_vote_annotations.xl results.xlsx + model.xlsx's (ka
generate_pie_charts.py	140	majority_vote_annotations.xl
cohens_kappa_pairwise_by_facet.py	155	GPT4/Ollama/QWEN.xlsx's
combined_pie_charts_figure.py	159	majority_vote_annotations.xl temporal_trend_normalized.xl
kappa_by_facet_batch_barchart.py	190	Human-results.xlsx (Batch 1/2/ xlsx's, majority_vote_annotati
kappa_combined_panels_figure.py	216	same as above (Batch 1/2/3, mode majority vote)
test_human_human_kappa_batch1_vs_batch2.py	219	Human-results.xlsx (Batch 1 &
cohens_kappa_batch1_heatmap.py	232	Human-results.xlsx (Batch 1), majority_vote_annotations.xl
cohens_kappa_batch2_heatmap.py	235	Human-results.xlsx (Batch 2), majority_vote_annotations.xl
overall_growth_changepoint_figure.py	242	temporal_trend_candidates_qu
test_mv_human_kappa_batch_comparisons.py	255	Human-results.xlsx (Batch 1/2/ xlsx's, majority_vote_annotati
temporal_trend_normalized.py	272	PubMed corpus.xlsx, majority_vote_annotations.xl temporal_trend_analysis_quarterly
majority_vote.py	344	GPT4/Ollama/QWEN.xlsx's

temporal_trend_analysis_quarterly.py	360	PubMed corpus.xlsx, majority_vote_annotations.xl
cohens_kappa_heatmap.py	369	Human-results.xlsx (Batch 3), majority_vote_annotations.xl
kappa_by_batch_barchart.py	449	Human-results.xlsx (Batch 1/2/ batch-anchored bars), model.xlsx' papers), majority_vote_annota

code/ · synthdata-review — generated from the repo's import graph and I/O calls, not from docstrings.