ICCV HONOLULU HAWAII
OCT 19-23, 2025

Effective Training Data Synthesis for Improving MLLM Chart Understanding

Yuwei Yang[1], Zeyu Zhang[1], Yunzhong Hou[1], Zhuowan Li[4], Gaowen Liu[3], Ali Payani[3],
Yuan-Sen Ting[2], Liang Zheng[1]
[1]Australian National University [2]Ohio State University
[3]Cisco [4]Johns Hopkins University

Paper ID: 3739

Chart Understanding Introduction

Charts

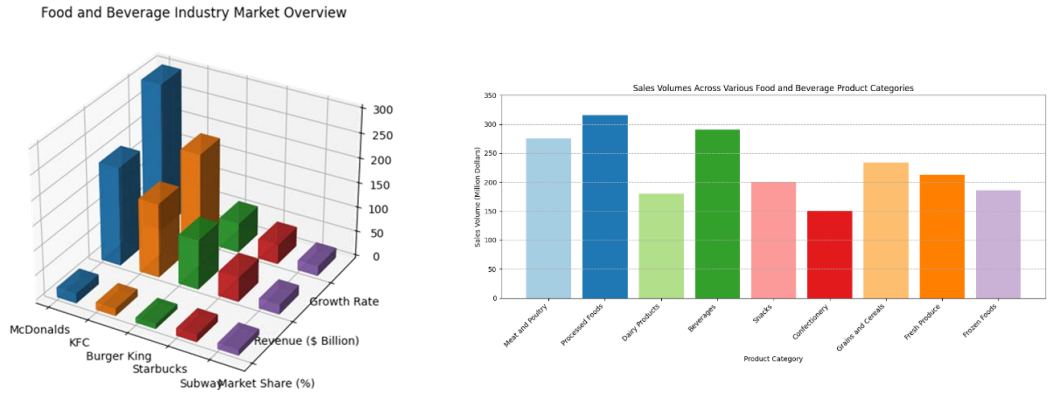


Chart QA

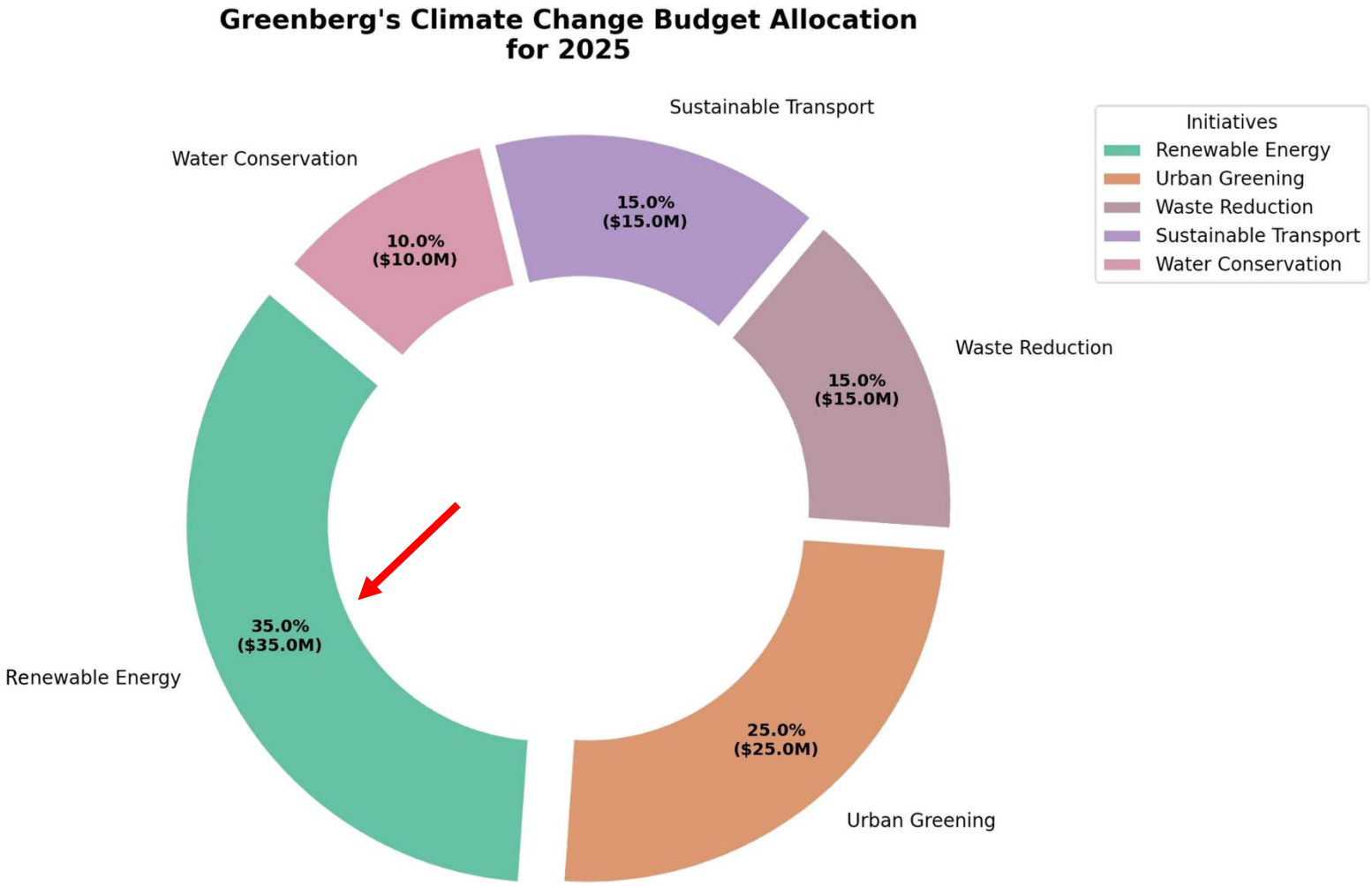
Chart Editing

Chart Description

Chart-to-Code

...

Chart Question Answering (QA) Task

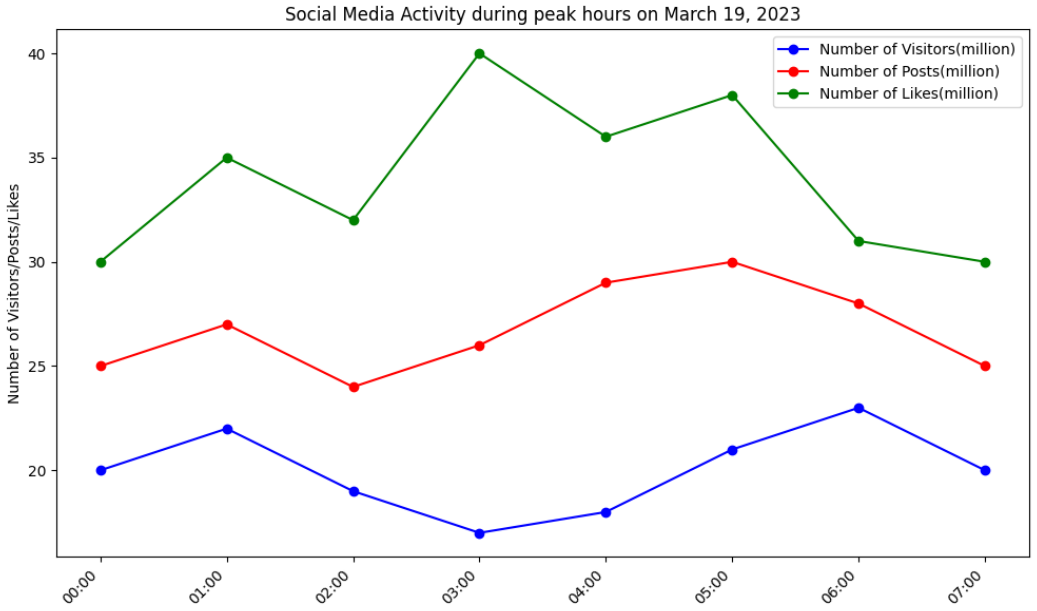


Question (Q): Which initiative has the highest percentage allocation in the chart?

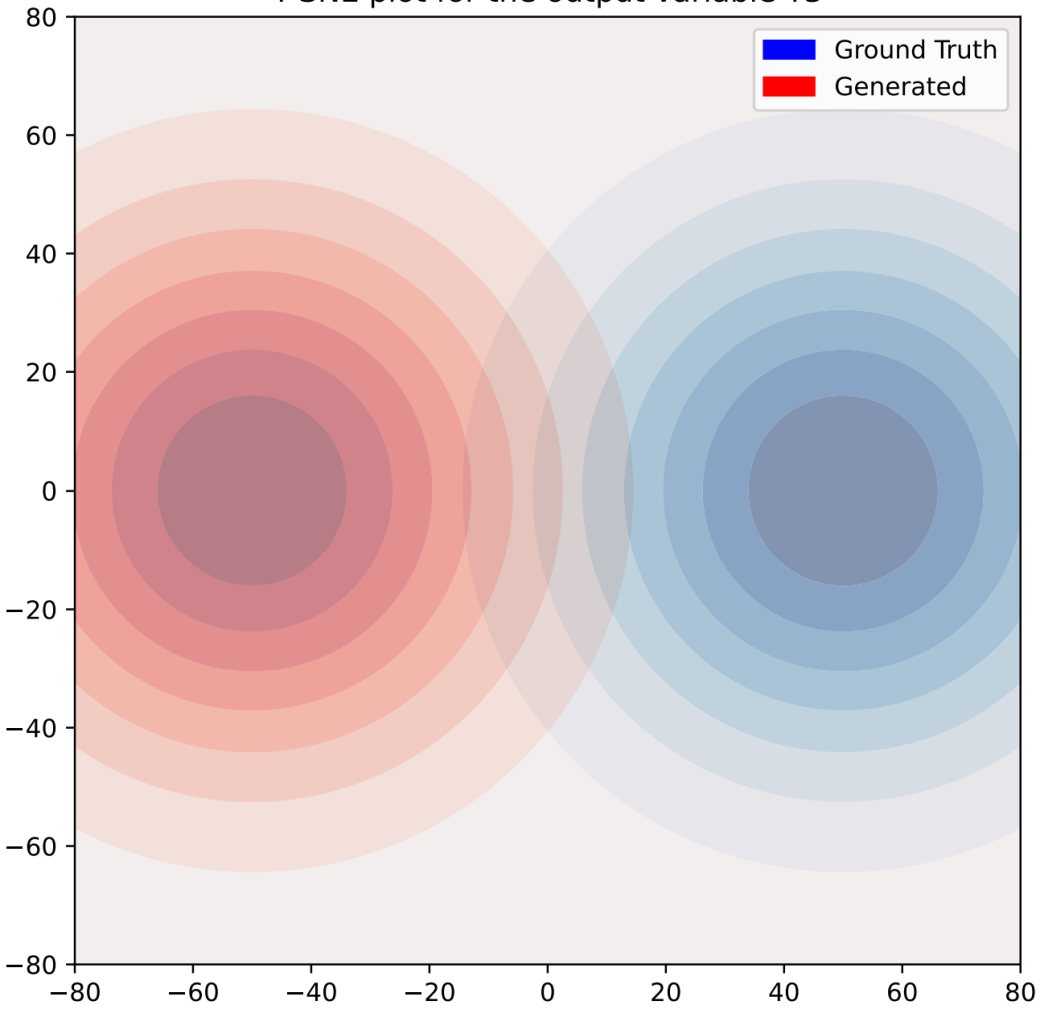
Answer (A): Renewable Energy

Previous Synthetic Chart QA Dataset: Limited Types

Common Type: Line Chart



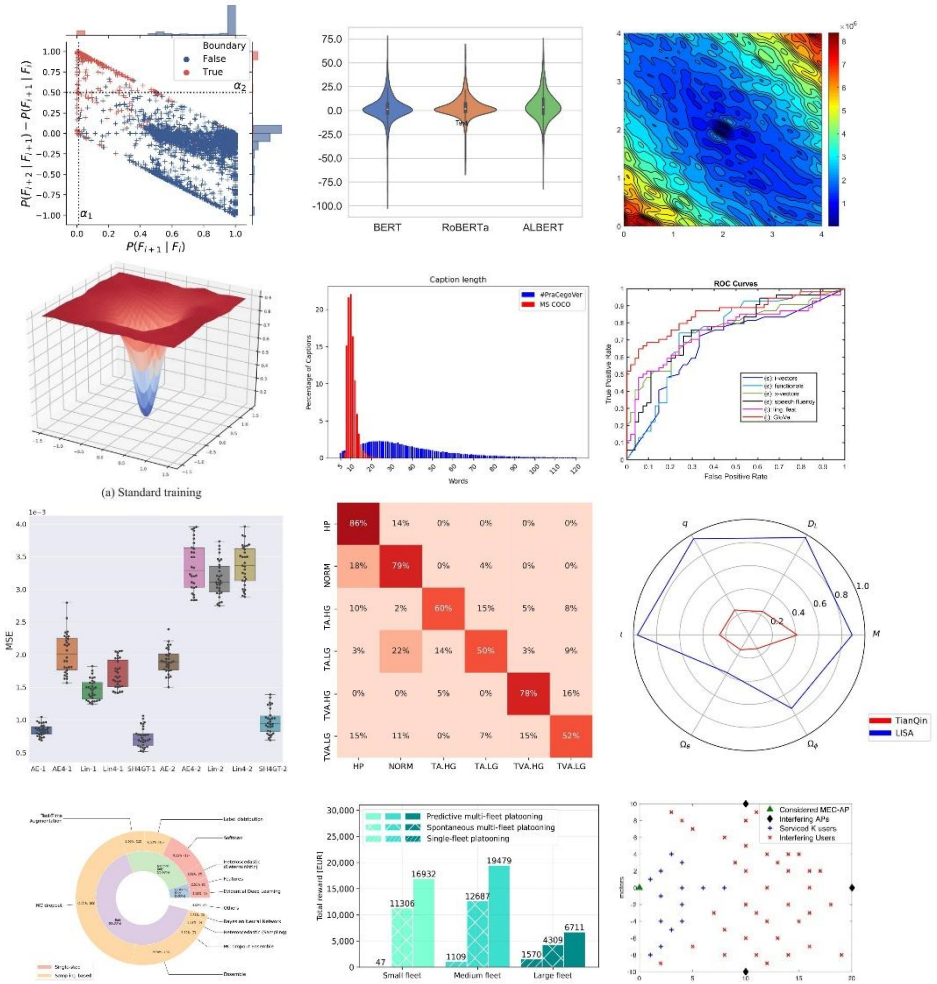
Novel Type: Contour Chart



Previous Synthetic Chart QA Dataset: Limited Types

Synthetic Chart Dataset

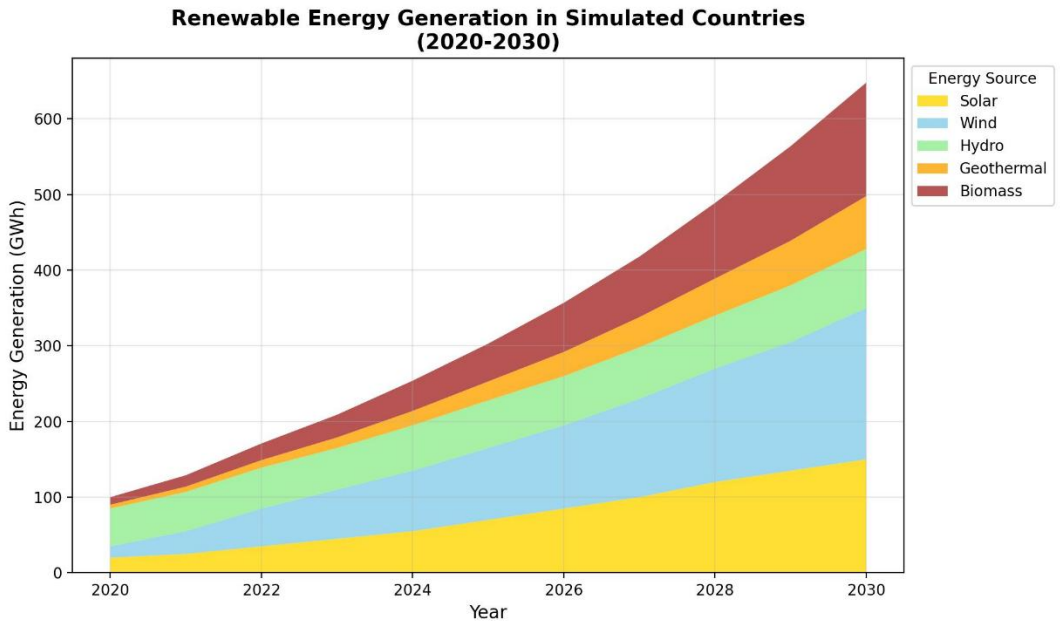
Real Chart Dataset



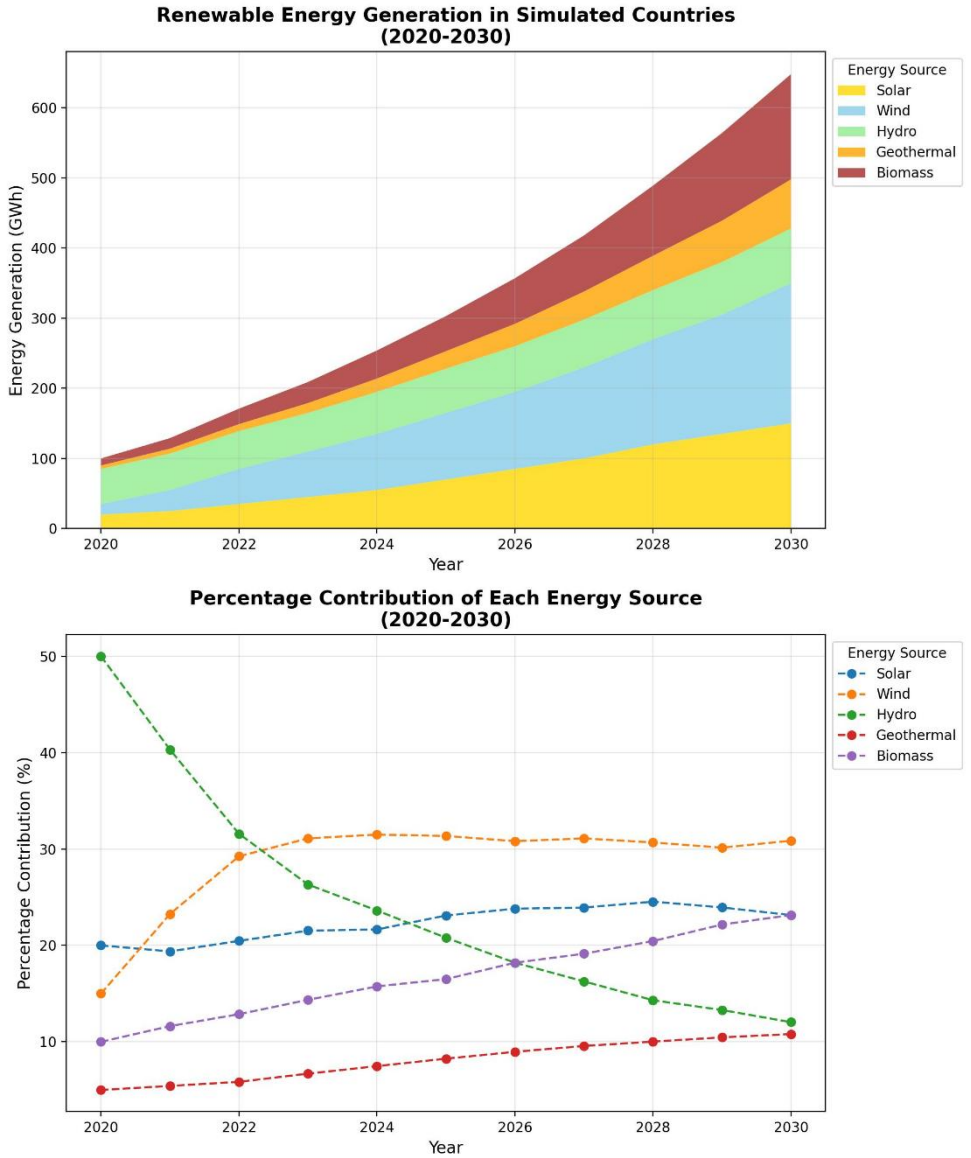
Various Chart Types

Previous Synthetic Chart QA Dataset: Single-Plot Only

Single-Plot



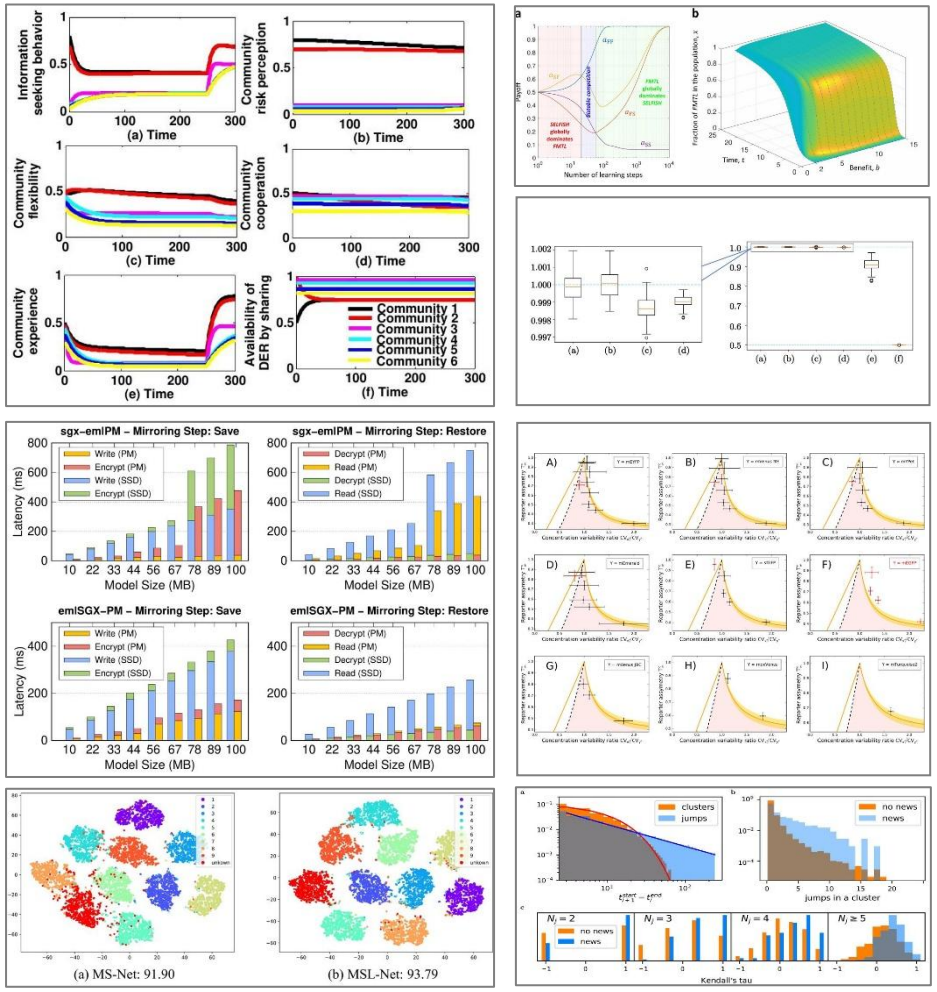
Combined Subplot



Previous Synthetic Chart QA Dataset: Single-Plot Only

Synthetic Chart Dataset

Real Chart Dataset



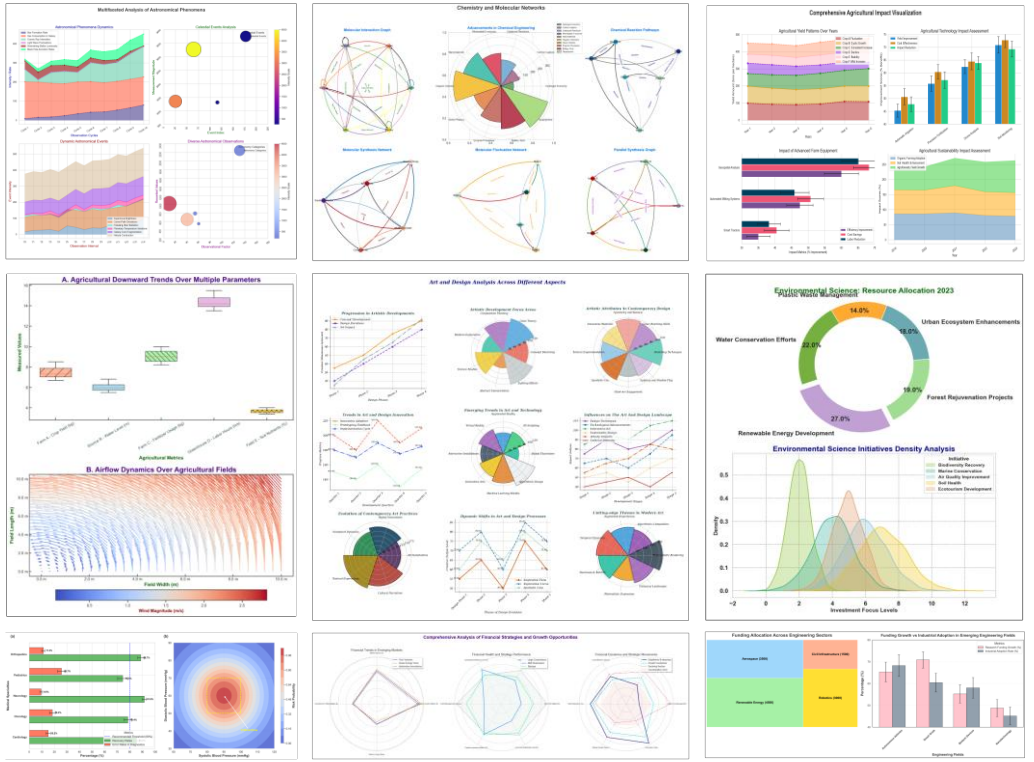
Complex Combined Subplot

●Our Synthetic Dataset: Expand Chart Types & Combined Subplot

29 Chart Types

…

250+ Types of Combined Subplot

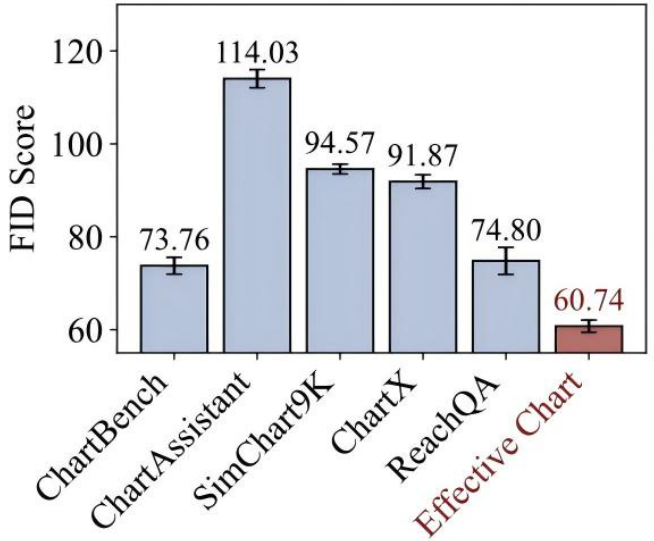


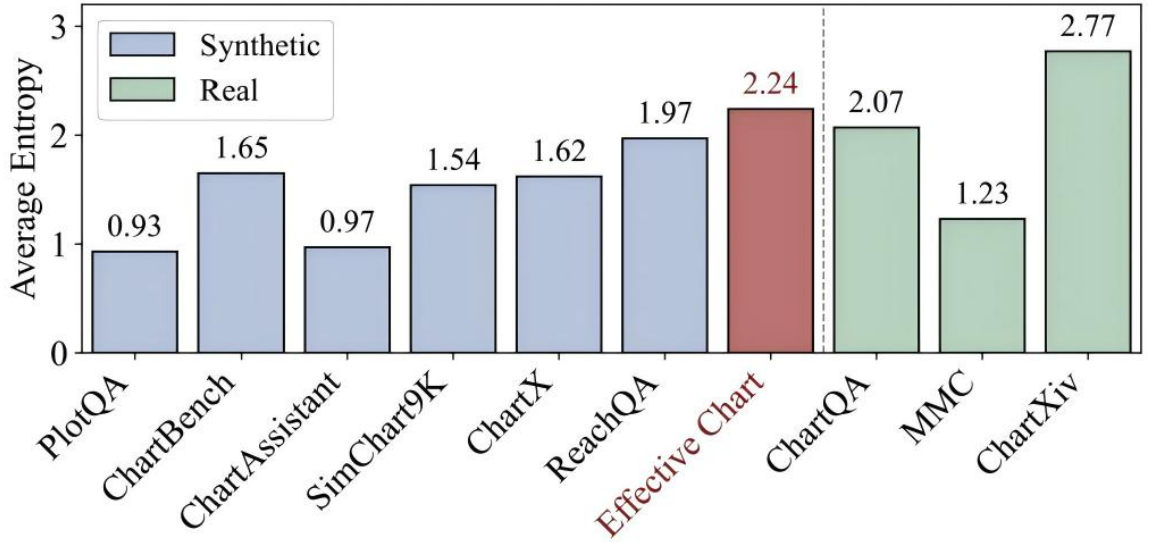
…

~10k Images, ~300k QAs

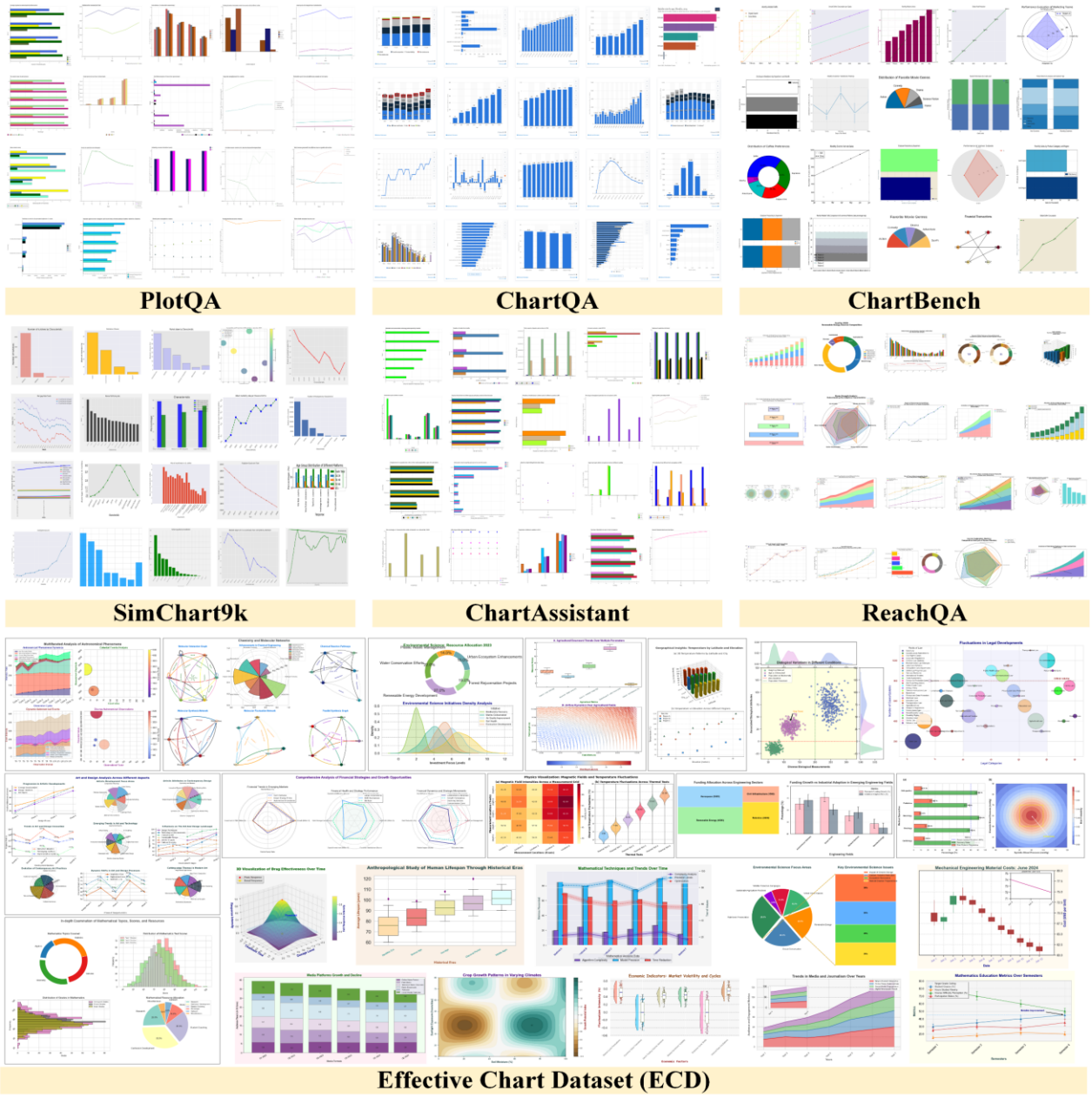
Quantitative Comparison with Other Datasets

Dataset	#Themes	#Chart types	#Type Combo.	#Images	Synthetic	Subplots	#QA Pairs	Training set
OpenCQA [24]	10	5	1	1.2k	×	×	1.2k	×
ChartX [47]	22	18	1	6k	✓	×	6k	×
CharXiv [44]	-	-	-	2.3k	×	✓	5k	×
PlotQA [39]	-	3	1	224k	✓	×	29M	✓
ChartQA [34]	15	3	1	21.9k	×	×	32.7k	✓
ChartLlama [14]	10	10	1	11k	✓	×	160k	✓
ChartBench [49]	9	9	4	68.7k	✓	×	618.5k	✓
ChartAssistant [38]	-	9	1	-	×	×	-	✓
SimChart9k [46]	-	3	1	9k	✓	×	9k	✓
MMC [30]	5	6	-	600k	×	✓	600k	✓
ReachQA [15]	-	10	-	3.7k	✓	✓	22k	✓
ECD (ours)	**25**	**29**	**252**	10.5k	✓	✓	321.5k	✓

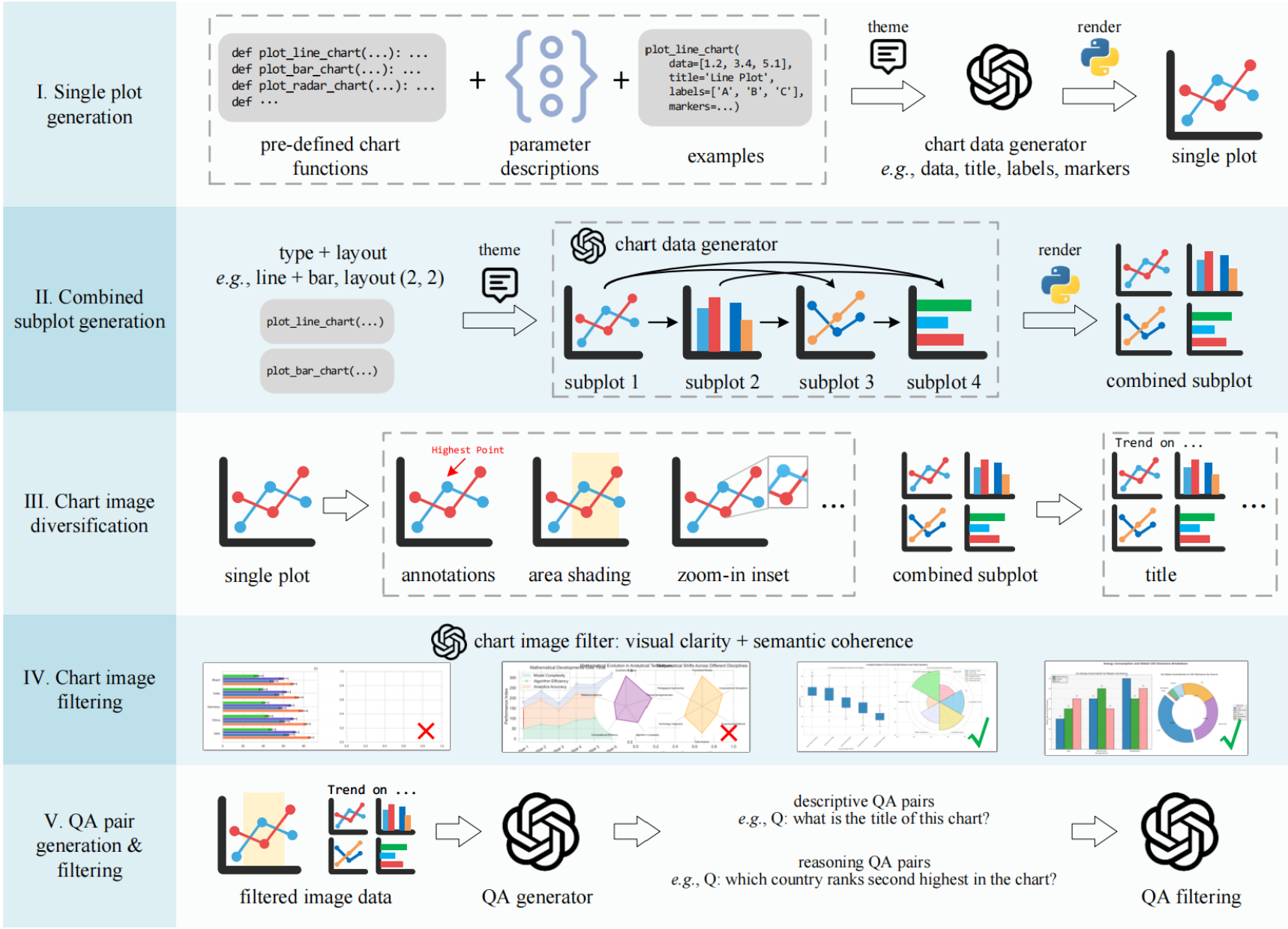




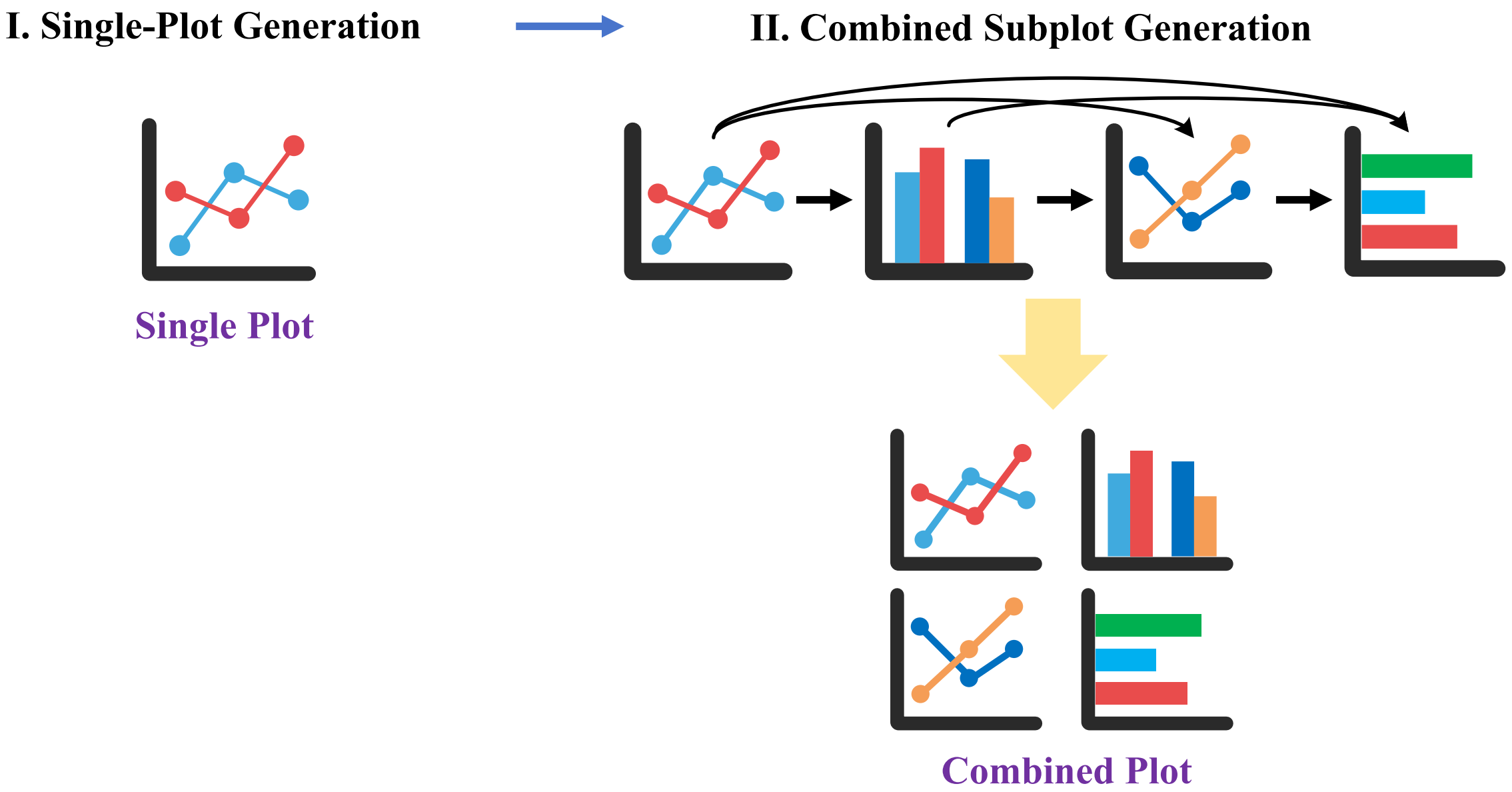
●Qualitative Comparison with Other Datasets



Our Data Synthesis Pipeline Overview

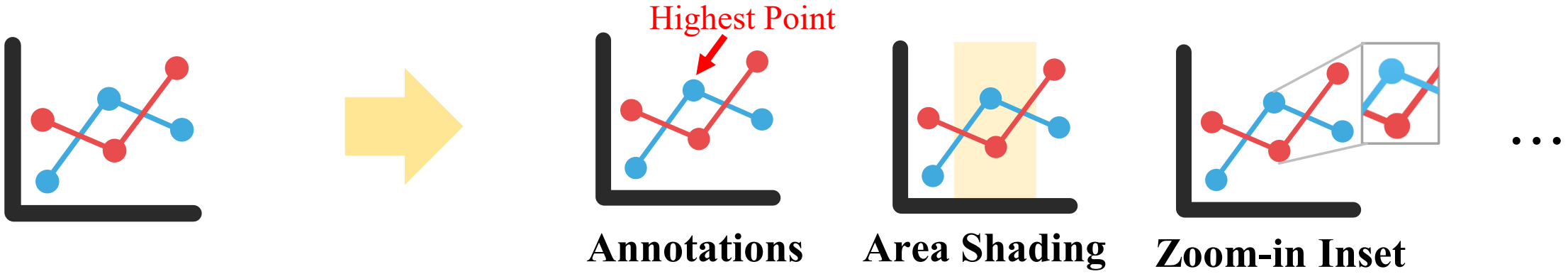


Our Data Synthesis Pipeline

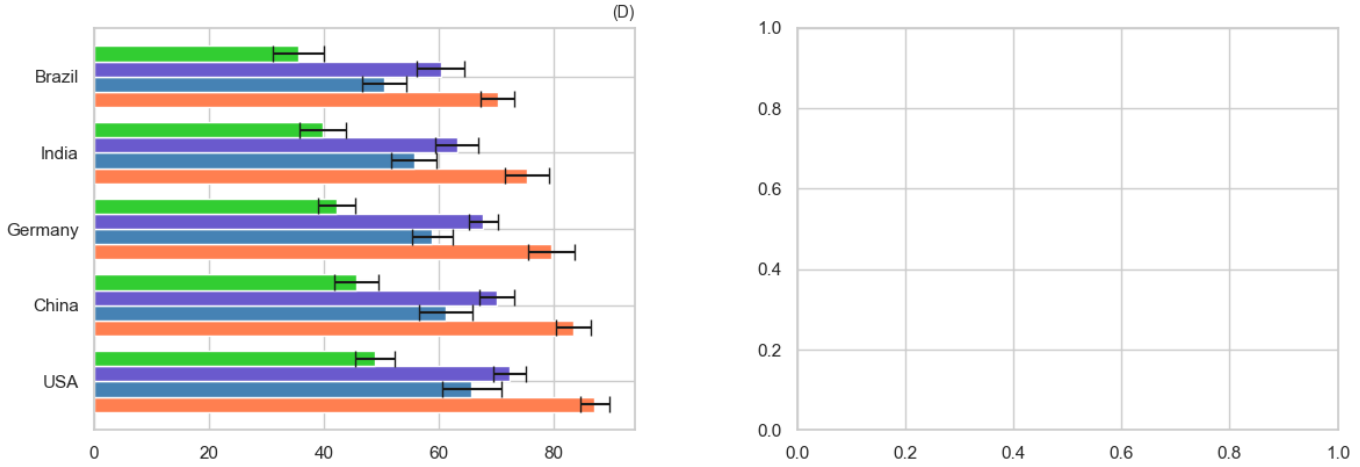


Our Data Synthesis Pipeline

III. Chart Image Divercification

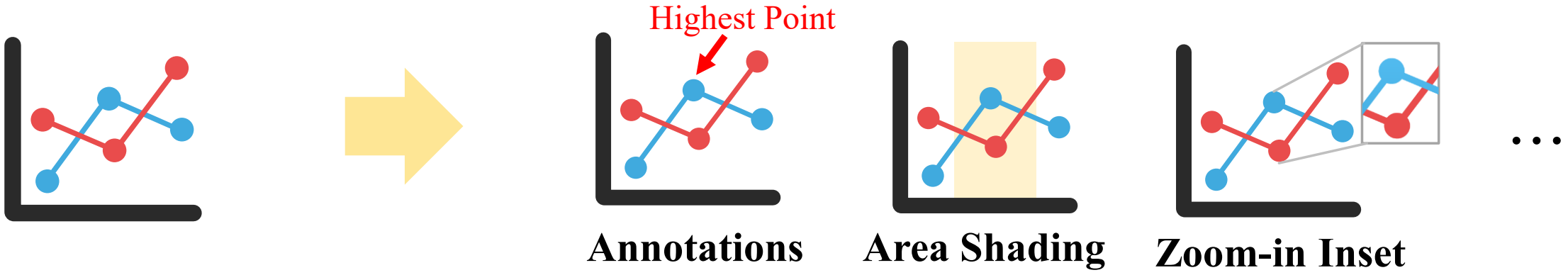


IV. Chart Image Filtering

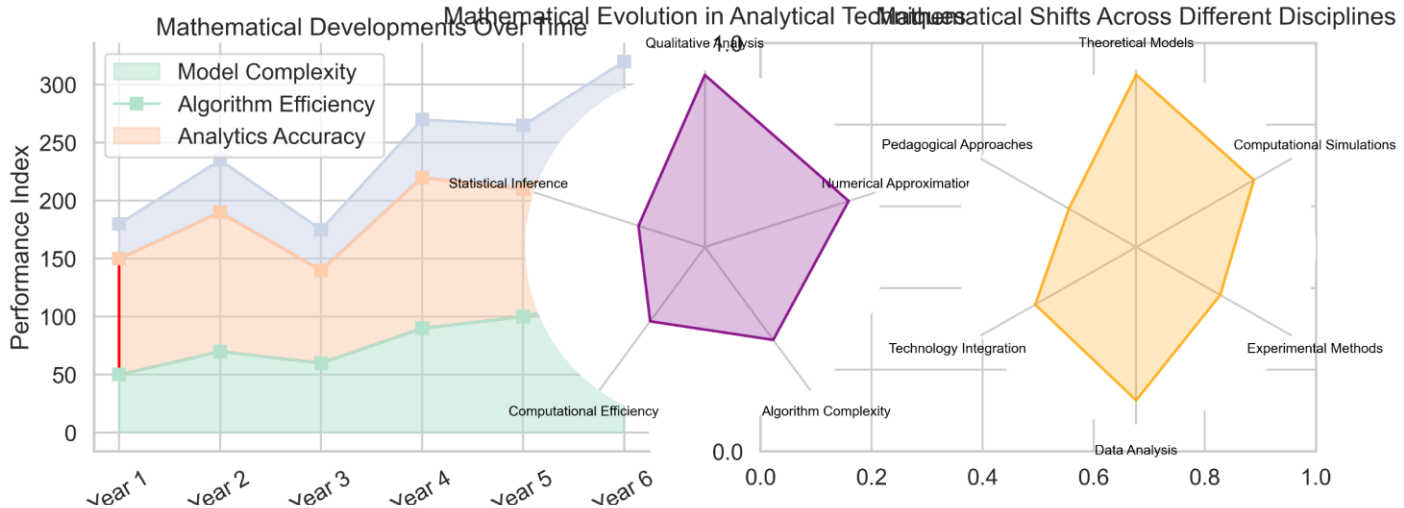


Our Data Synthesis Pipeline

III. Chart Image Divercification

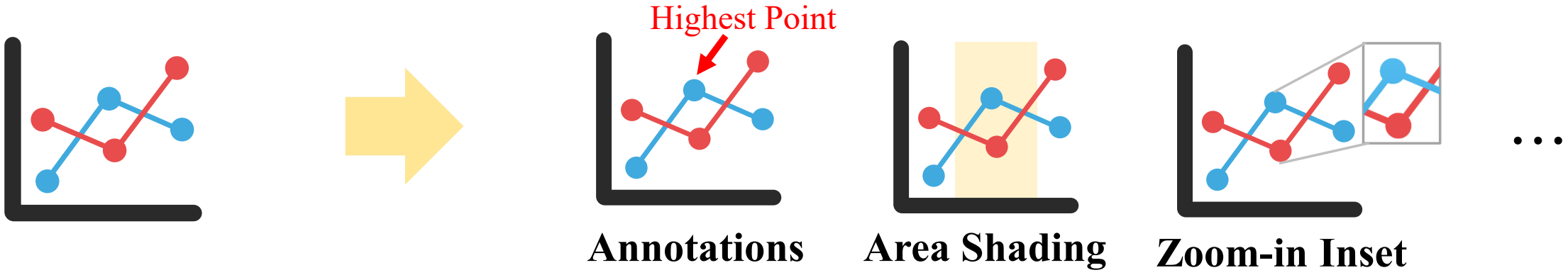


IV. Chart Image Filtering

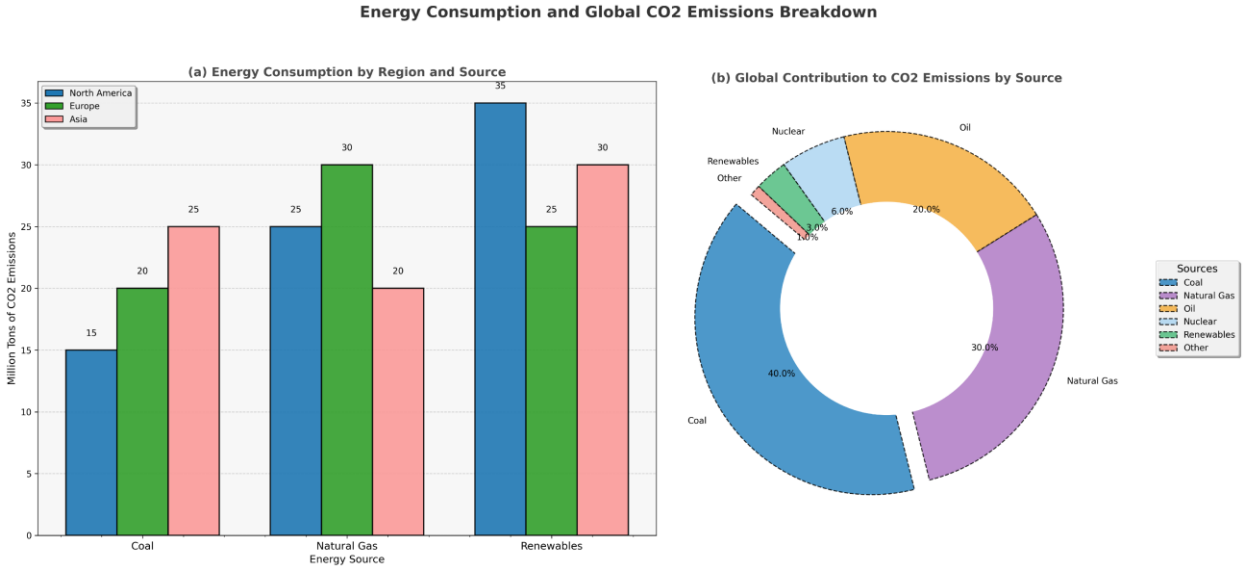


●Our Data Synthesis Pipeline

III. Chart Image Divercification



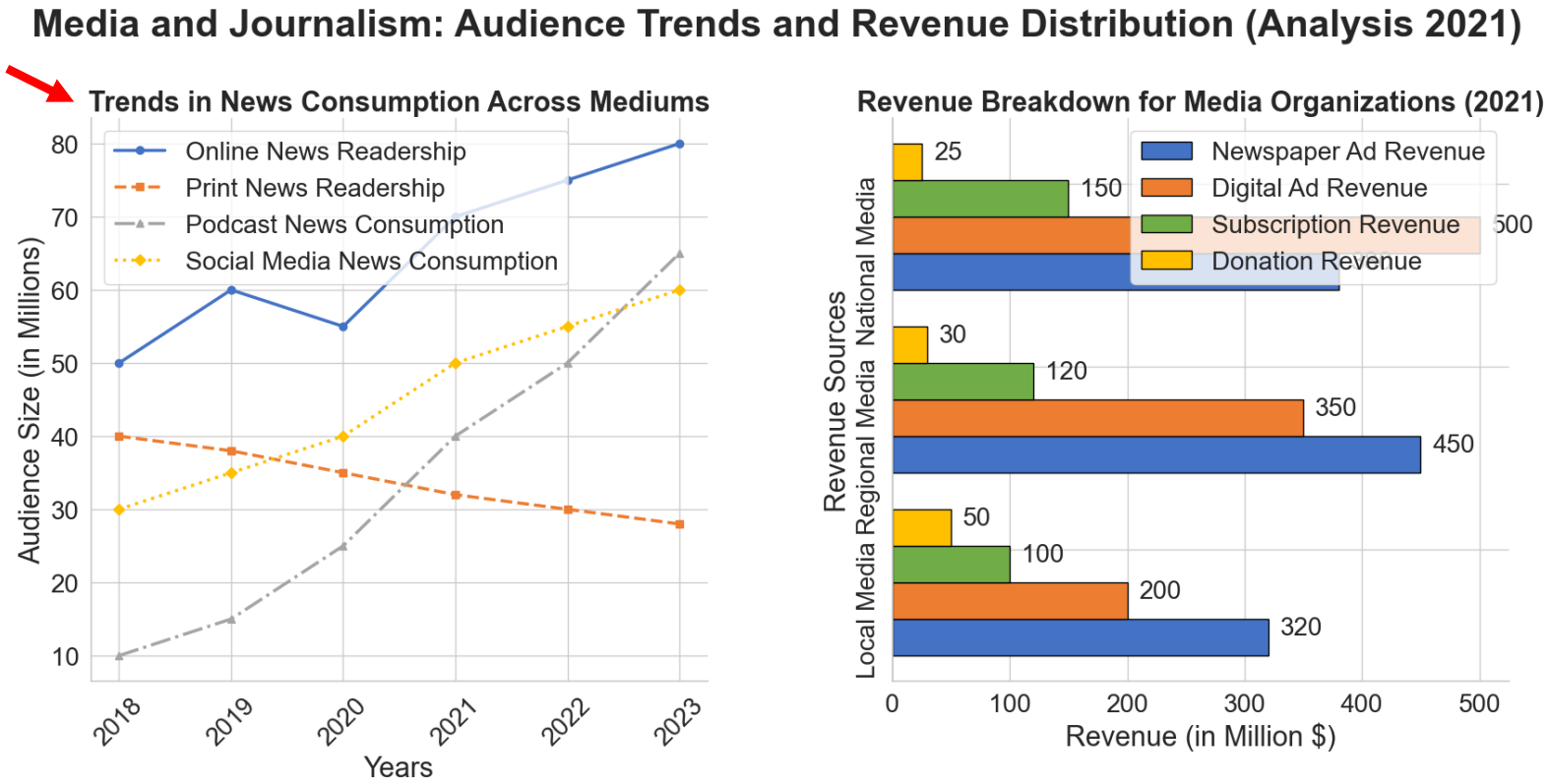
IV. Chart Image Filtering



√

●Our Data Synthesis Pipeline

IV. Chart QA Generation



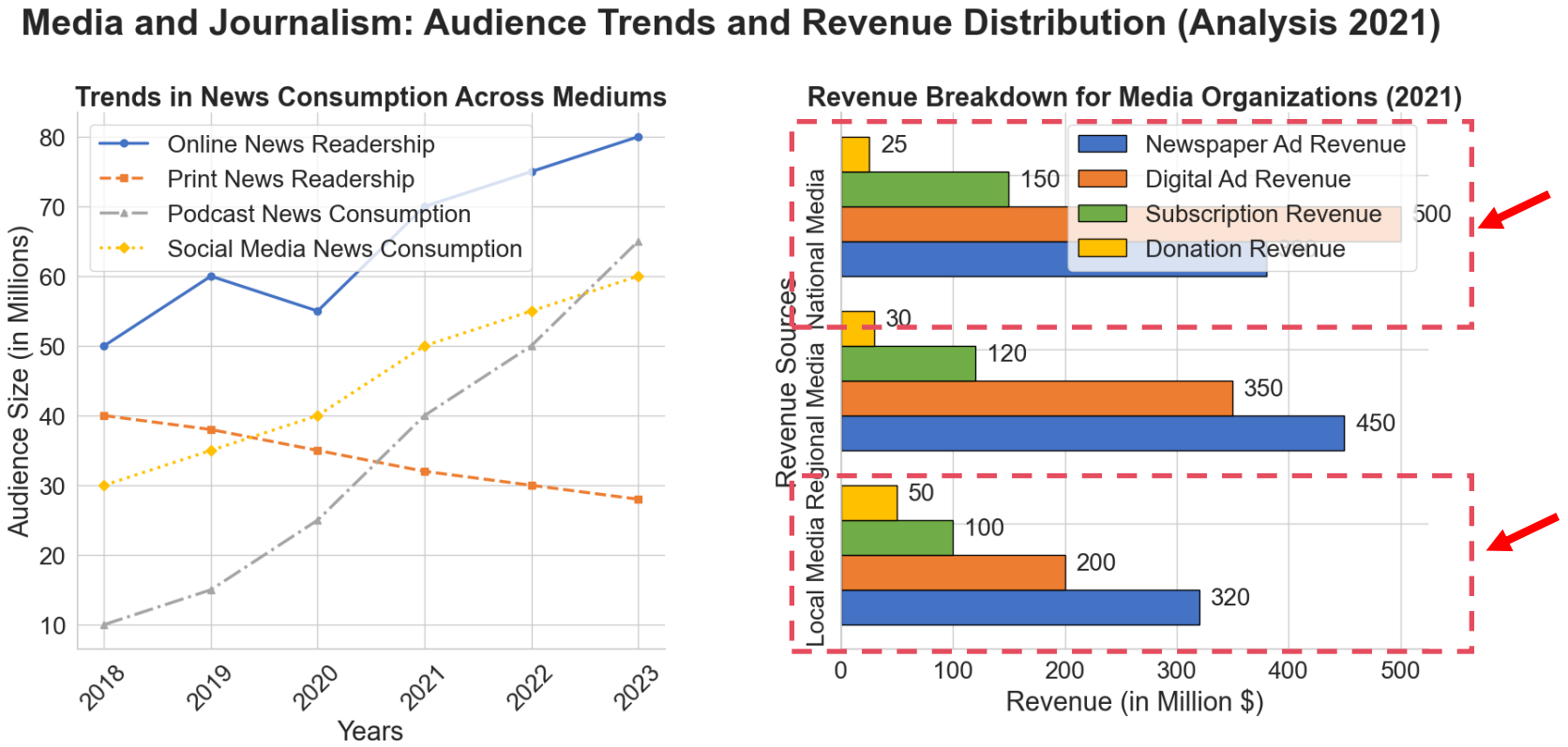
Descriptive QA:

Question (Q) : What is the title of the left subplot?

Answer (A): The title of the left subplot is 'Trends in News Consumption Across Mediums'.

●Our Data Synthesis Pipeline

IV. Chart QA Generation

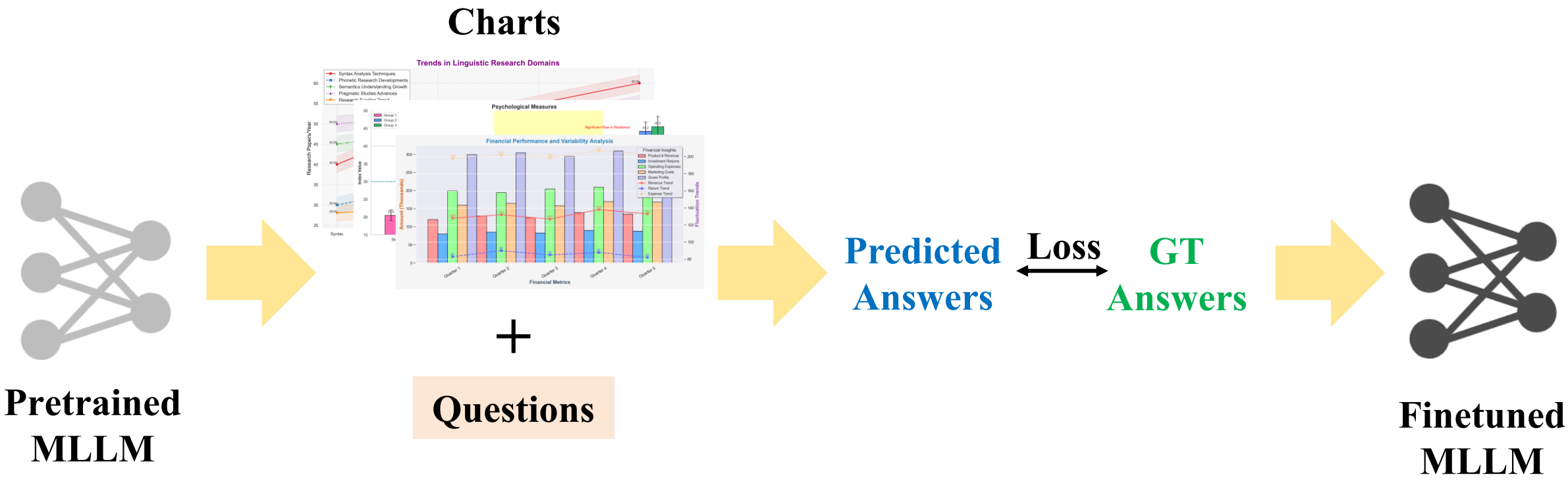


Reasoning QA:

Question (Q) : Which revenue source shows the largest difference between Local Media and National Media?

Answer (A): The differences for each source are: Newspaper Ad Revenue = 380 - 320 = 60M, Digital Ad Revenue = 500 - 200 = 300M, Subscription Revenue = 150 - 100 = 50M, Donation Revenue = 25 - 50 = -25M. Largest difference is for Digital Ad Revenue with 300M. Answer: Digital Ad Revenue

●Model Training on Our Dataset – Supervised-Finetuning (SFT)



●Our ECDBench Construction

- 📊 **Charts**: 1,224
  - **Single-plot**: 364
  - **Multi-subplot**: 860
    - 2 types: 457
    - 3 types: 403
- 🖥️ **Avg. Resolution**: 1378 × 968 px
- 🤖 Constructed via **GPT-4o + human refinement**
- ✍️ **Q&A Pairs**: 2,448
  - 1 Descriptive + 1 Reasoning QA per image

Model Performance Comparison (%)

Model	Reasoning	Descriptive	Average
o4-mini	**57.03**	**77.45**	**67.24**
o3	56.13	74.51	65.32
Gemini-2.5-Pro	44.36	76.88	60.62
o1	40.52	74.18	57.35
Claude-4-Sonnet	44.20	69.36	56.78
Claude-3.7-Sonnet	43.38	69.61	56.50
Claude-3.5-Sonnet	41.99	68.14	55.07
Qwen2.5-VL-72B	38.81	68.46	53.64
GPT-4o	35.62	70.18	52.90
GPT-4o-mini	24.26	57.27	40.77
Qwen2.5-VL-32B	24.92	53.92	39.42
Qwen2.5-VL-7B	19.04	57.35	38.19
Random (GPT-4o)	4.58	1.63	3.10

Model Evaluation on Chart Benchmark

Models	CharXiv			ChartQA	ReachQA			ChartBench			ChartX	ECDBench		
	Rea.	Des.	Avg.	QA	Rea.	Recog.	Avg.	Binary	NQA	Avg.	QA	Rea.	Des.	Avg.
Baselines														
Human	80.50	92.10	89.78	-	65.10	84.60	74.85	-	-	-	-	-	-	-
Random (GPT-4o) [4]	10.80	19.85	18.04	30.04	8.20	13.30	10.75	-	22.73	-	11.46	4.58	1.63	3.10
Proprietary multimodal large language models														
GPT-4o mini [4]	34.10	74.92	66.76	77.52	27.20	53.50	40.35	-	34.93	-	44.36	24.26	57.27	40.77
GPT-4o [4]	47.10	84.45	76.98	85.70	39.70	66.80	53.25	-	52.88	-	58.33	35.62	70.18	52.90
Claude 3.5 Sonnet [5]	60.20	84.30	79.48	90.80	51.70	74.30	63.00	-	48.29	-	42.71	41.99	68.14	55.07
Open-source multimodal large language models														
LLaVA-Next-Llama3-8B [22]	19.70	38.90	35.06	64.56	8.00	23.30	15.65	61.18	38.29	58.64	27.69	4.74	17.16	10.95
+ ECD	21.80	59.05	51.60	68.64	12.70	37.50	25.10	64.21	47.05	62.30	46.61	16.50	46.65	31.58
MiniCPM-V2.6 [45]	25.40	52.65	47.20	76.96	14.50	48.10	31.30	75.67	52.33	73.08	47.31	15.15	39.95	27.53
+ ECD	26.90	61.90	54.91	77.28	17.50	48.40	32.95	70.77	52.81	68.77	50.43	18.14	52.21	35.17
Phi-3-Vision [2]	31.50	60.52	54.72	81.92	28.40	62.80	45.60	71.92	62.71	70.90	67.53	21.65	41.18	31.41
+ ECD	33.40	68.00	61.08	84.88	32.80	64.50	48.65	73.49	63.76	72.41	71.44	29.49	59.31	44.40
Qwen2.5-VL-7B [6]	41.20	66.40	61.36	83.04	29.70	71.90	50.80	80.99	67.81	79.53	67.80	19.04	57.35	38.19
+ ECD	40.20	74.20	67.40	85.32	36.20	70.80	53.50	79.35	70.86	78.41	70.83	35.38	66.34	50.86

Training set	Test set					
	CharXiv	ChartQA	ReachQA	ChartBench	ChartX	ECDBench
Baseline	35.06	64.56	15.65	58.64	27.69	10.95
ChartQA	35.16	**68.92**	15.00	61.41	31.51	13.11
ChartBench	32.86	61.56	18.35	**75.13**	37.33	10.99
ReachQA	30.68	64.50	24.35	57.59	39.24	13.48
ECD	**51.60**	68.64	**25.10**	62.30	**46.61**	**31.58**

Training Set Comparison:
Fine-tuning
LLaVA-Next-Llama3-8B shows
ECD > others

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Effective Training Data Synthesis for Improving MLLM Chart Understanding

Project: *https://effective-chart-dataset-synthesis.github.io/*

THANKS