

Scores Heatmap

Dataset: evals.parquet

Slide1

```
from inspect_viz import Data
from inspect_viz.plot import plot
from inspect_viz.mark import cell, text

evals_data = Data.from_file("evals.parquet")

plot(
    cell(
        evals_data,
        x="task_name",
        y="model",
        fill="score_headline_value",
        tip=True,
        inset=1,
        sort={
            "y": {"value": "fill", "reduce": "sum", "reverse": True},
            "x": {"value": "fill", "reduce": "sum", "reverse": False},
        },
    ),
    text(
        evals_data,
        x="task_name",
        y="model",
        text="score_headline_value",
        fill="white",
        styles={"font_weight": 600},
    ),
    padding=0,
    color_scheme="viridis",
    height=250,
    margin_left=150,
    x_label=None,
    y_label=None
)
```

openai/o3 –	0.35	0.359	0.933	0.793	0.851
openai/o4-mini –	0.41	0.403	0.833	0.793	0.829
:/claude-sonnet-4-20250514 –	0.35	0.674	0.5	0.667	0.834
pic/claude-3-7-sonnet-latest –	0.275	0.516	0.267	0.672	0.806
	cybench	swe_bench	aime2024	gpqa_diamond	mmlu_pro