

Exploratory Data Analysis (EDA) Report

COVID Vaccine Statewise Dataset

Dataset file: covid_vaccine_statewise.csv

Rows: 7845 Columns: 24

Numeric cols (sample): Total Doses Administered, Sessions, Sites , First Dose Administered, Second

Categorical cols (sample): State

Datetime cols: Updated On

This report includes: variable types, summary statistics, missing-value handling (median & KNN), outlier treatment (IQR capping), categorical encoding, correlation heatmap, and visualizations (histogram, boxplot, violin, bubble, 2D density).

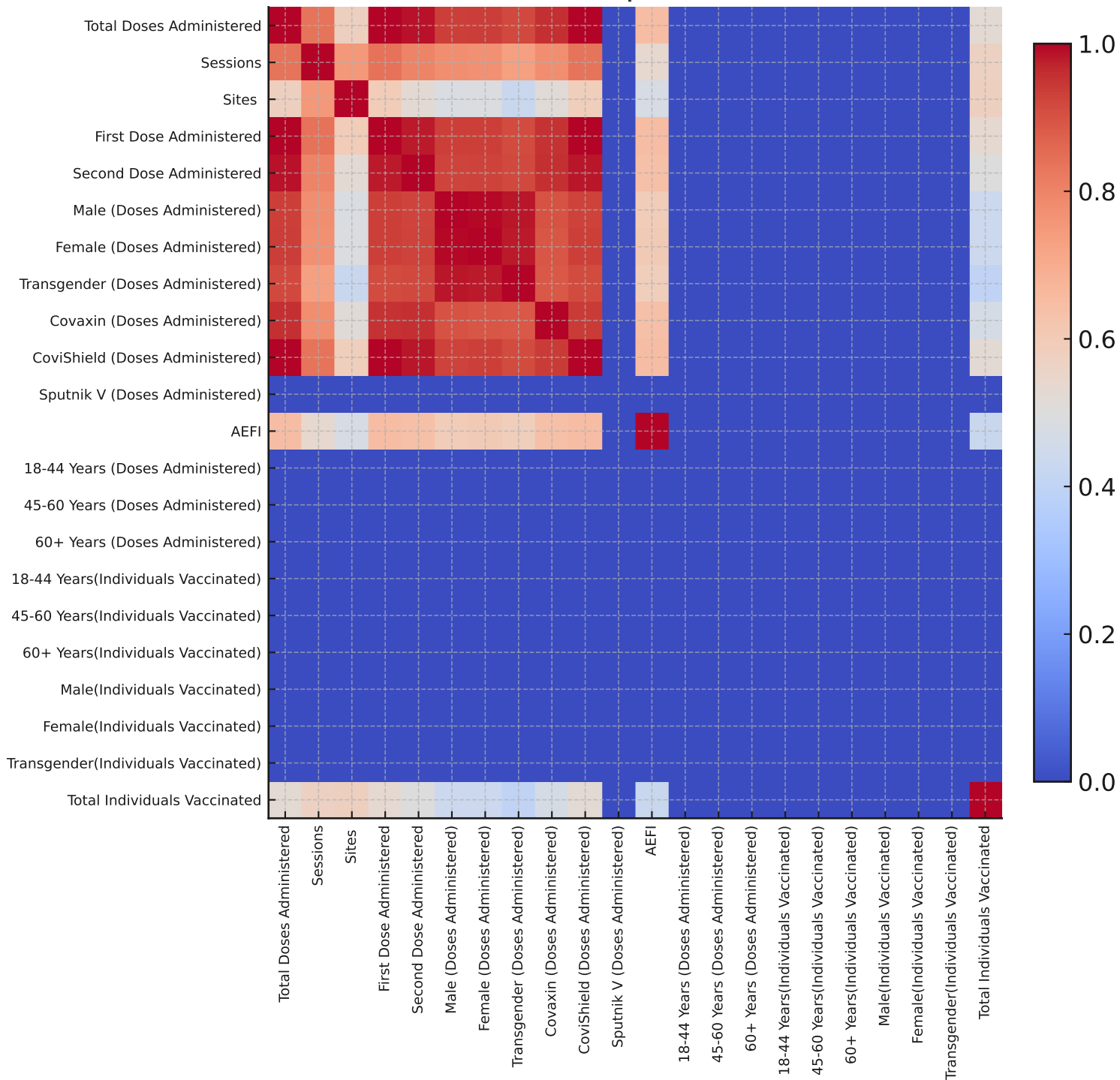
Numeric Summary (first rows)

	count	mean	std	min	25%	50%	75%	max
stered	7621.0	9188170.544	87461801.17	7.0	135657.0	818202.0	6625243.0	513228400.0
	7621.0	479235.799	1911511.193	0.0	6004.0	45470.0	342869.0	35010311.0
	7621.0	2282.872	7275.974	0.0	69.0	597.0	1708.0	73933.0
tered	7621.0	7414415.3	29952087.78	7.0	116632.0	661459.0	5387805.0	400150406.0
nistered	7621.0	1773755.244	7570382.402	0.0	12831.0	138818.0	1166434.0	113077994.0
nistered)	7461.0	8620156.0111	7379382.717	0.0	56555.0	389785.0	2735777.0	270163622.0
nistered)	7461.0	3168416.361	5153103.672	2.0	52107.0	334238.0	2561513.0	239518609.0
s Administered)	7461.0	1162.978	5931.354	0.0	8.0	113.0	800.0	98275.0
dministered)	7621.0	1044669.322	4452258.87	0.0	0.0	11851.0	757930.0	62367416.0
Administered)	7621.0	8126552.9243	2984142.544	7.0	133134.0	756736.0	6007817.0	446825051.0
administered)	2995.0	9655.571	43882.536	0.0	0.0	0.0	2519.0	588039.0
	5438.0	1139.403	3454.608	0.0	109.25	294.0	808.0	26542.0

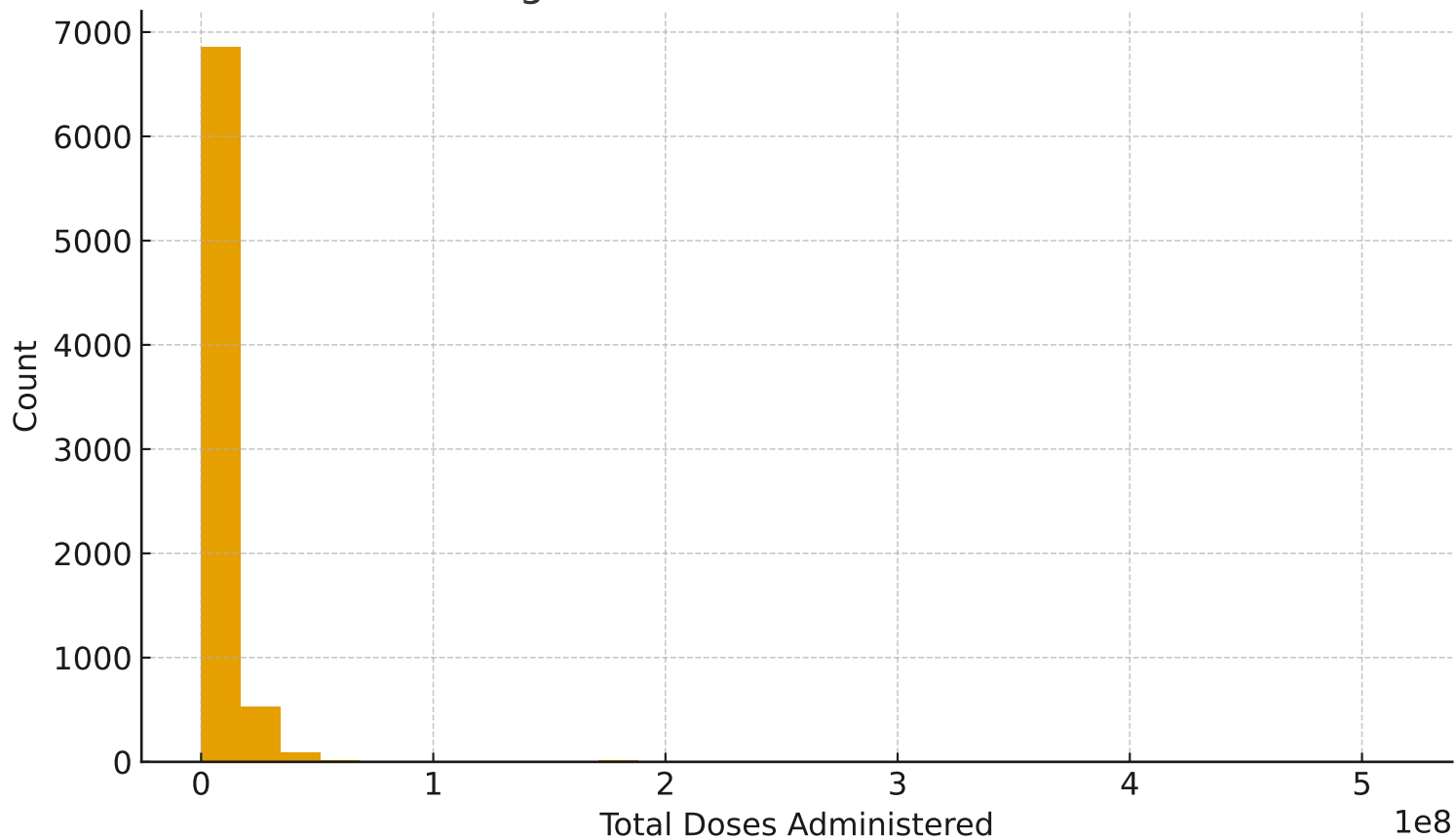
Missing values: before and after (top rows)

	missing_before	missing_after_stat	missing_after_knn
	0	0	0
	0	0	0
ered	224	0	0
	224	0	0
	224	0	0
red	224	0	0
stered	224	0	0
tered)	384	0	0
nistered)	384	0	0
Administered)	384	0	0
nistered)	224	0	0
dministered)	224	0	0
nistered)	4850	0	0
	2407	0	0
Administered)	6143	0	0
Administered)	6143	0	0
nistered)	6143	0	0
als Vaccinated)	4112	0	0
als Vaccinated)	4111	0	0
s Vaccinated)	4111	0	0
inated)	7685	0	0
accinated)	7685	0	0
als Vaccinated)	7685	0	0
inated	1926	0	0

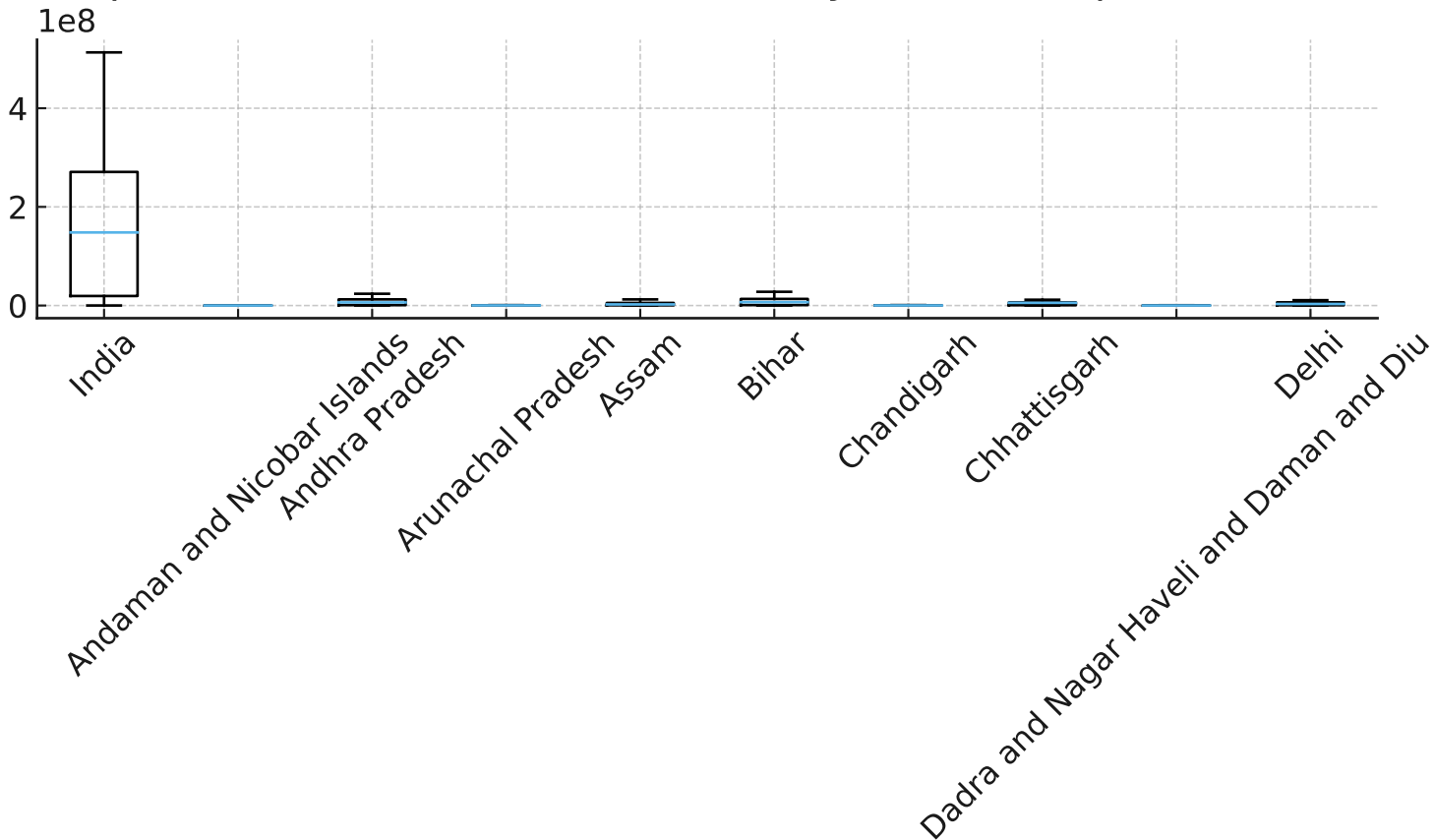
Correlation heatmap (numeric features)



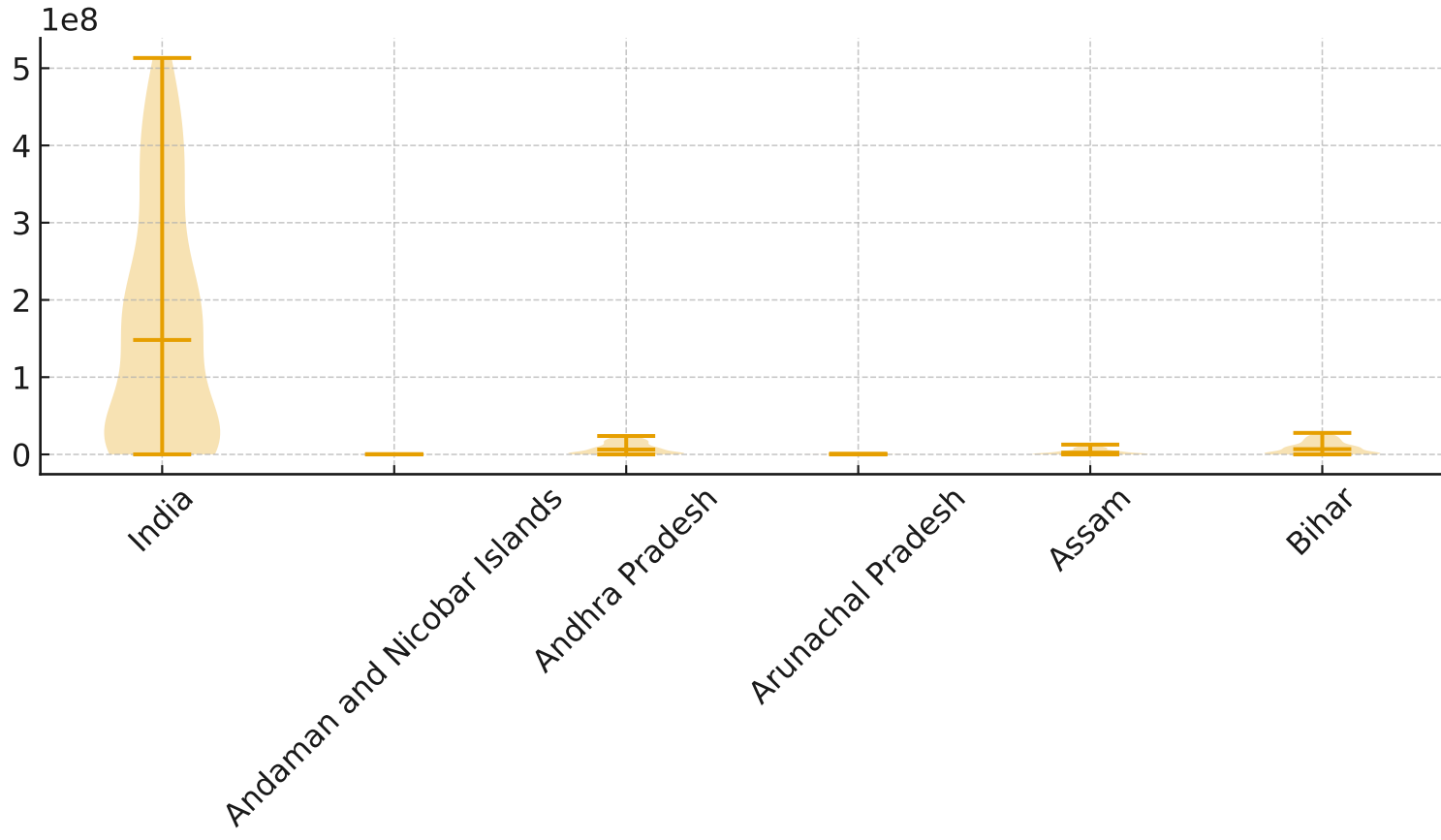
Histogram: Total Doses Administered



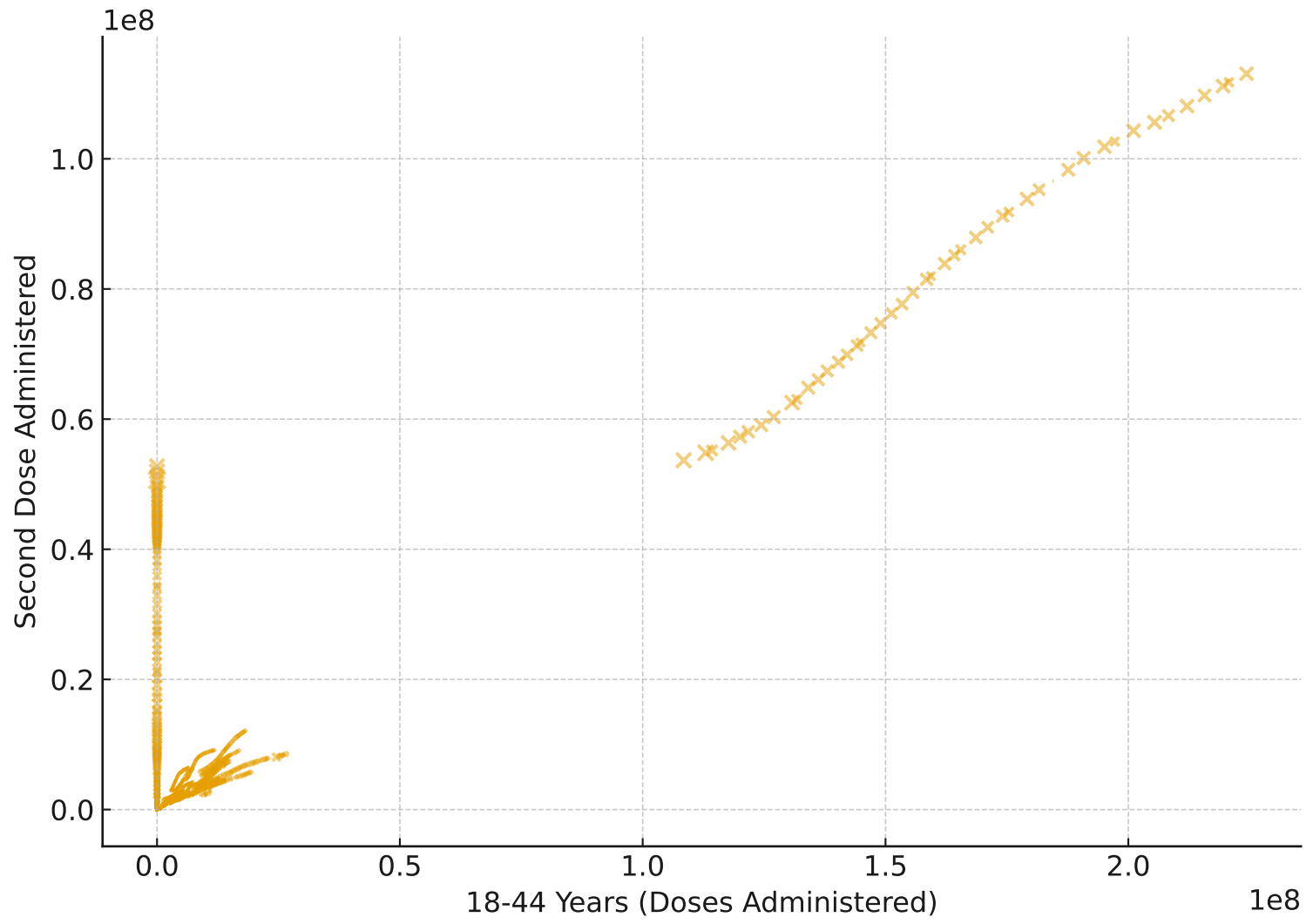
Boxplot of Total Doses Administered by State (sample 10 states)



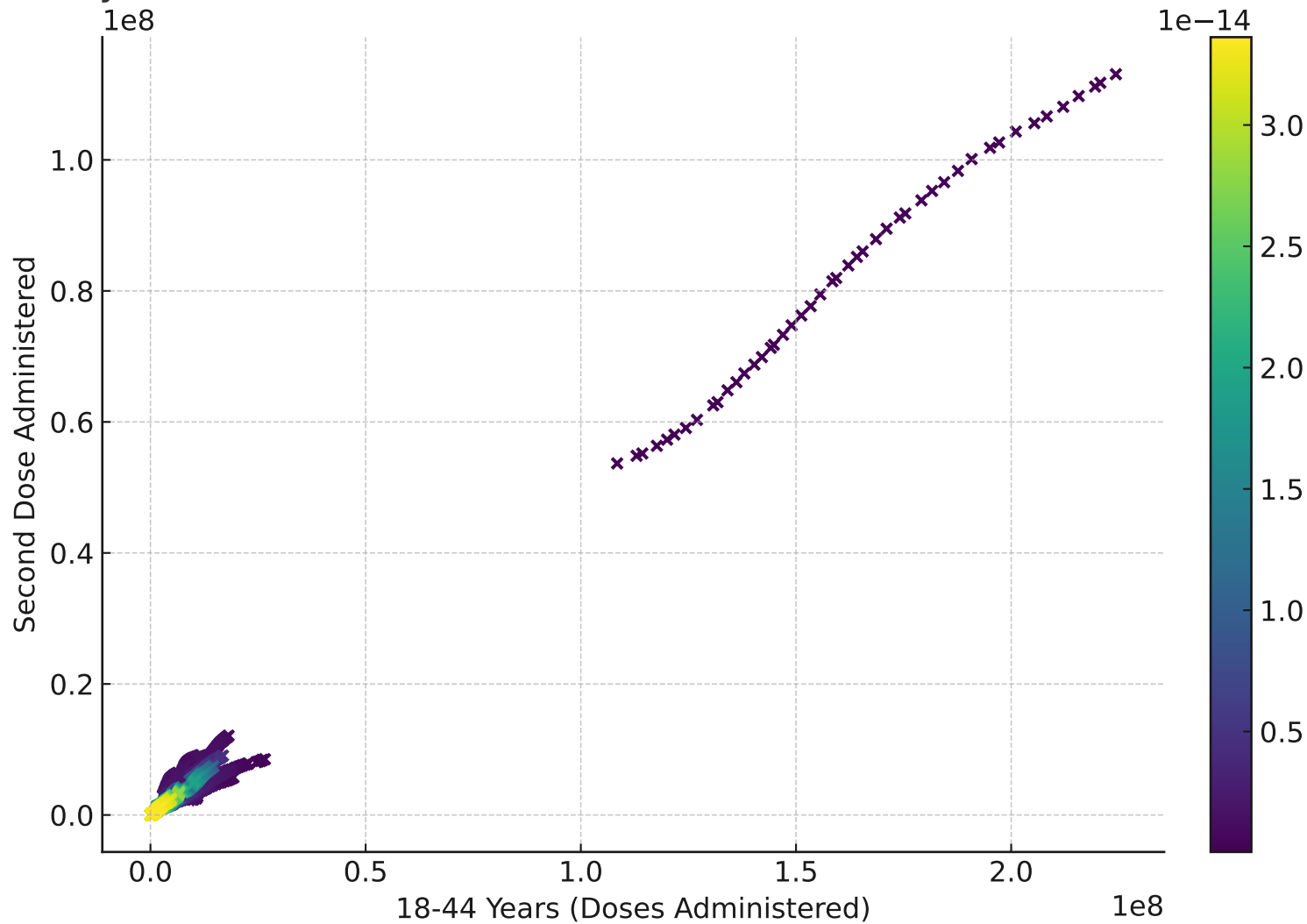
Violin plot: Total Doses Administered by State (sample categories)



Plot: 18-44 Years (Doses Administered) vs Second Dose Administered (size



density: 18-44 Years (Doses Administered) vs Second Dose Administered



Code snippets and methodological notes

Representative code snippets used in this EDA:\n1) Median imputation for numeric columns:\nfor col in nume