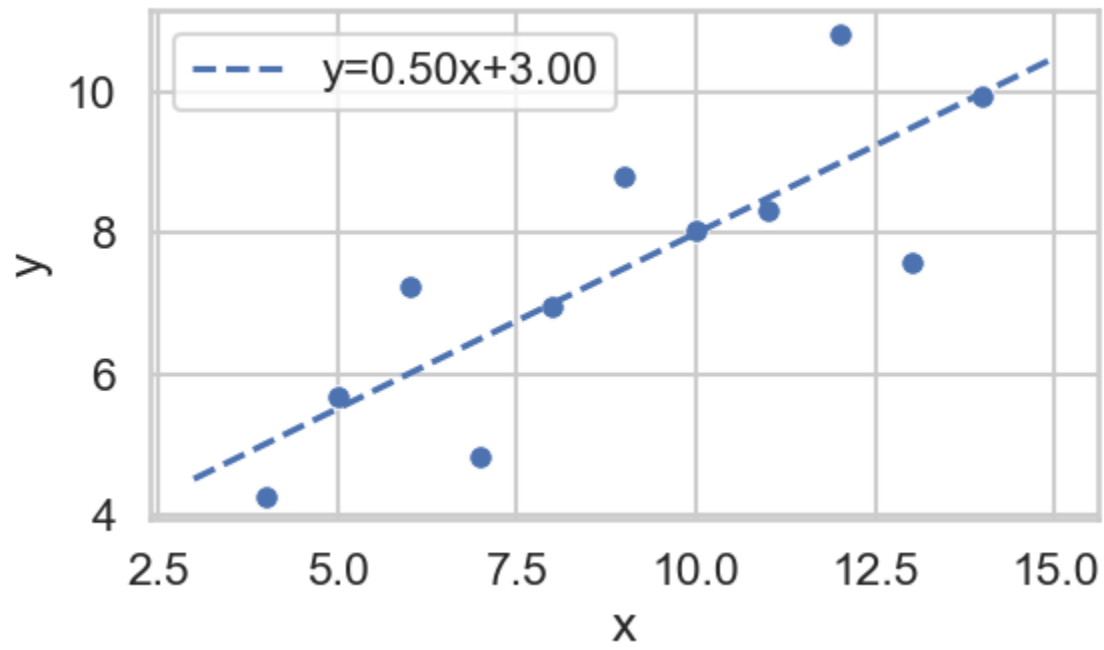Dataset I — Scatter + OLS

y=0.50x+3.00

Dataset III — Scatter + OLS

y=0.50x+3.00

Dataset IV — Scatter + OLS

y=-0.19x+9.23
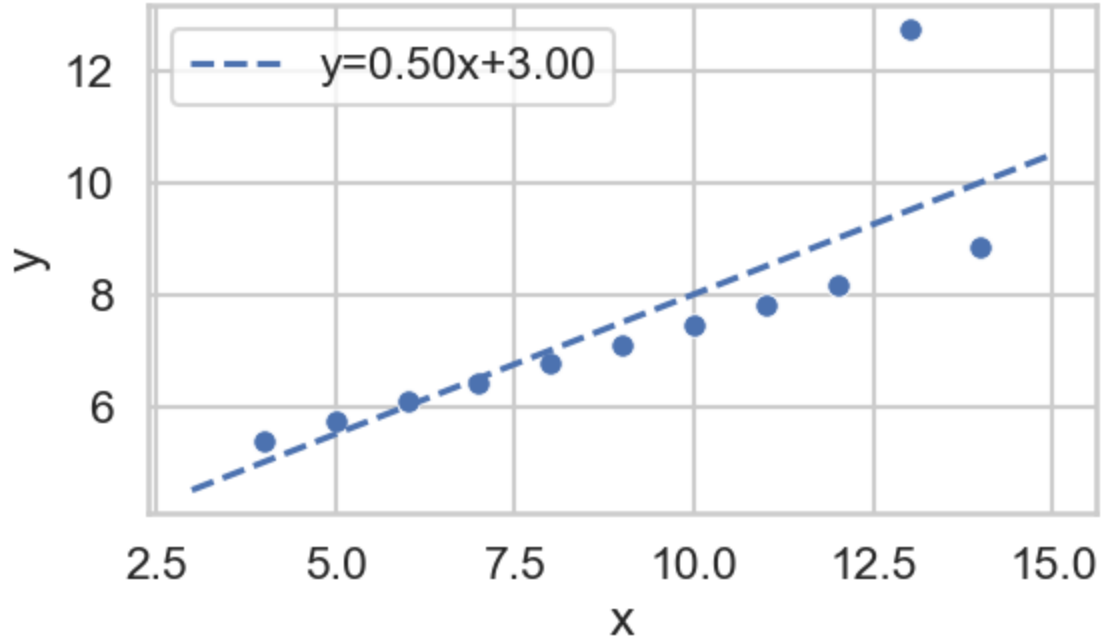
Dataset I — Residuals

Dataset II — Residuals
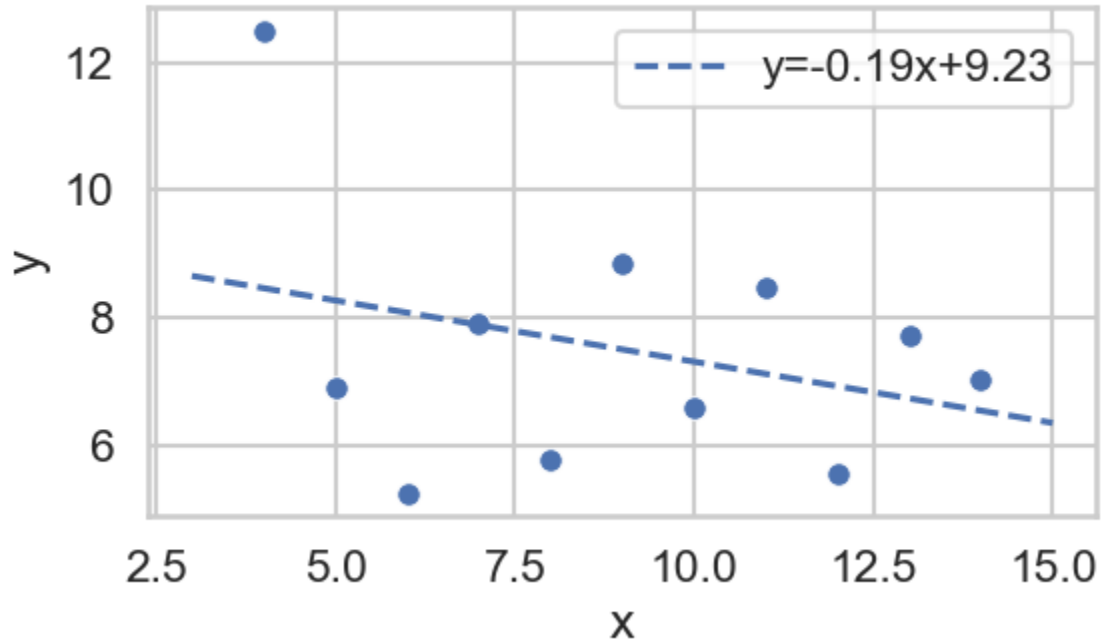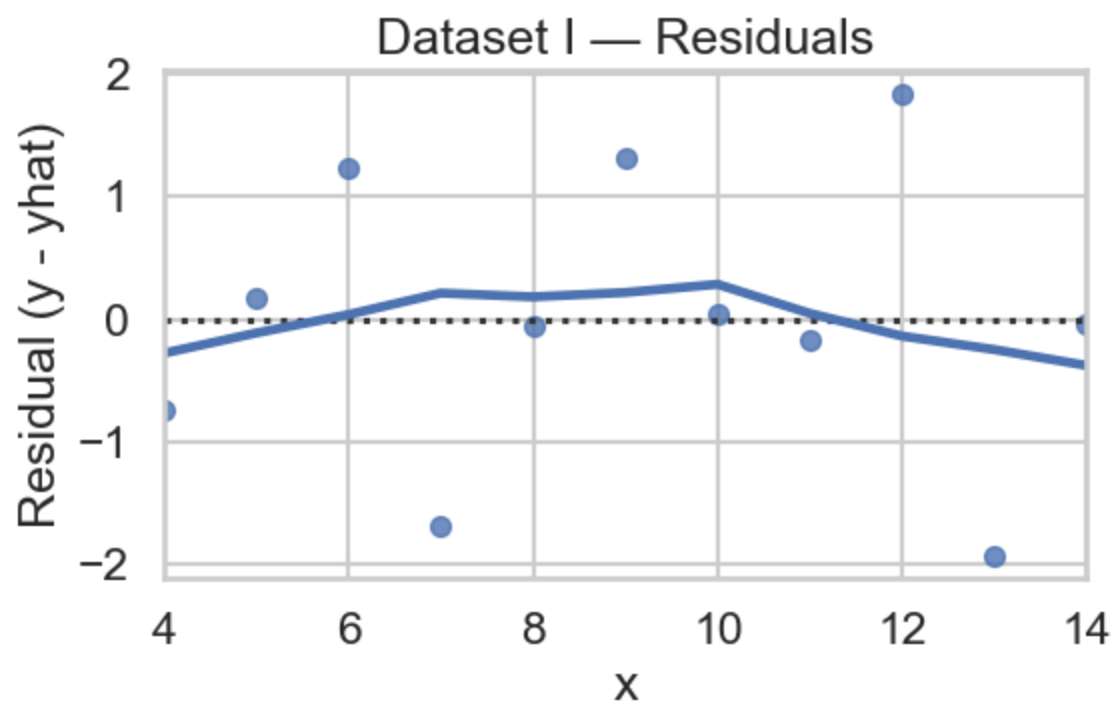
Dataset III — Residuals

Dataset IV — Residuals

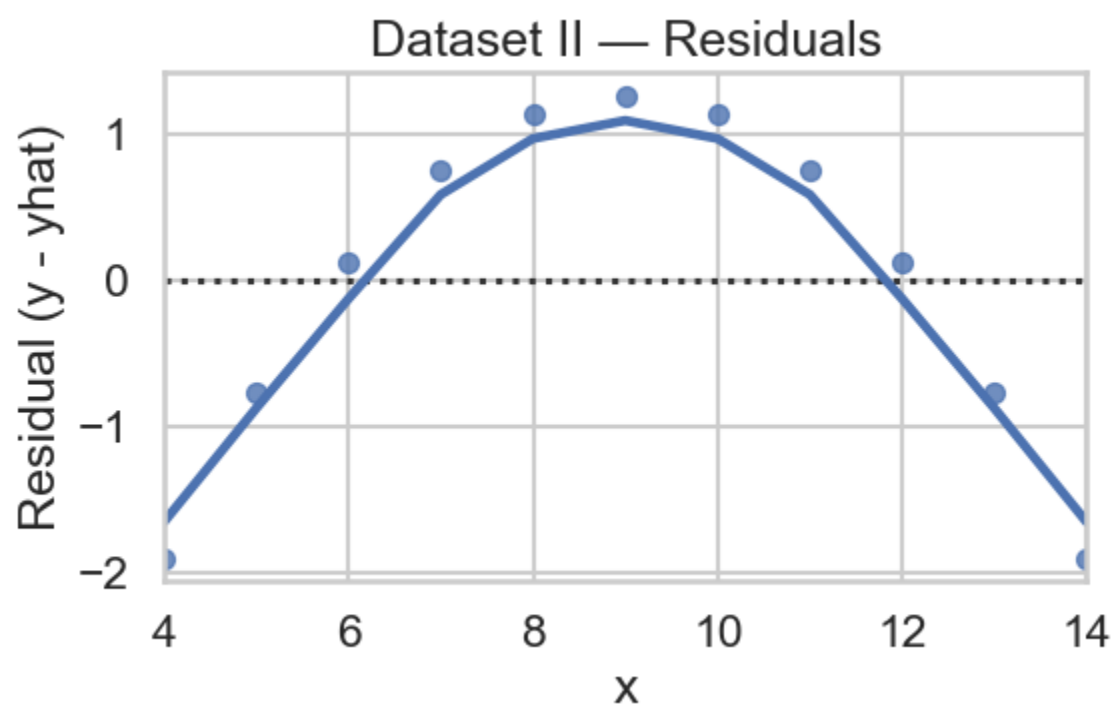Boxplot of x per dataset

Boxplot of y per dataset

Dataset II — Scatter + OLS

y=0.50x+3.00
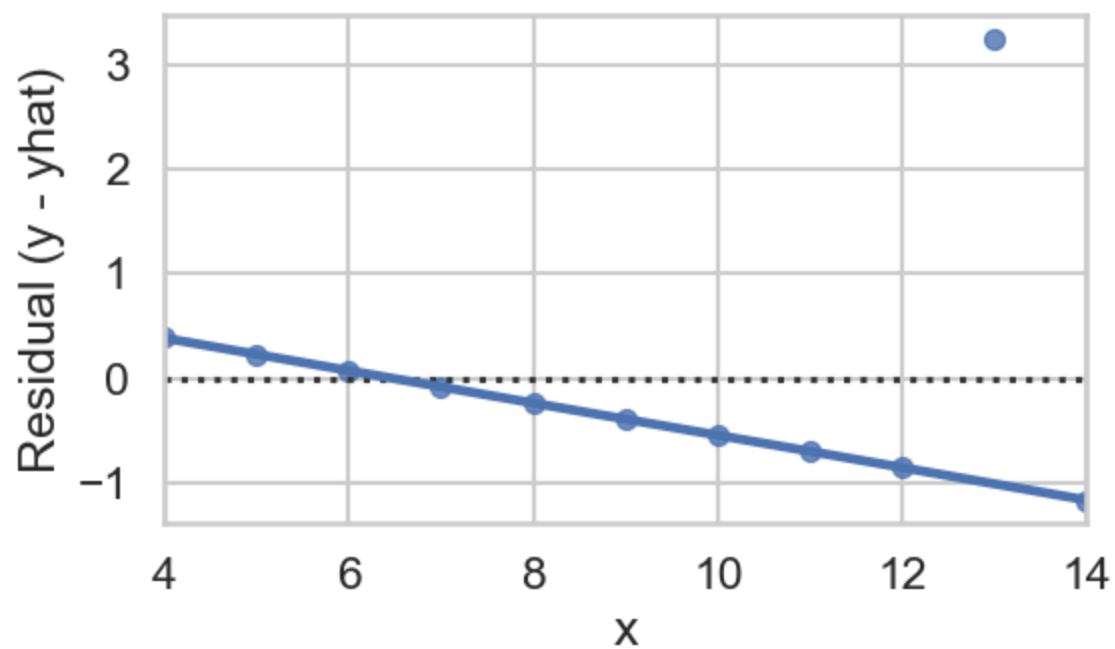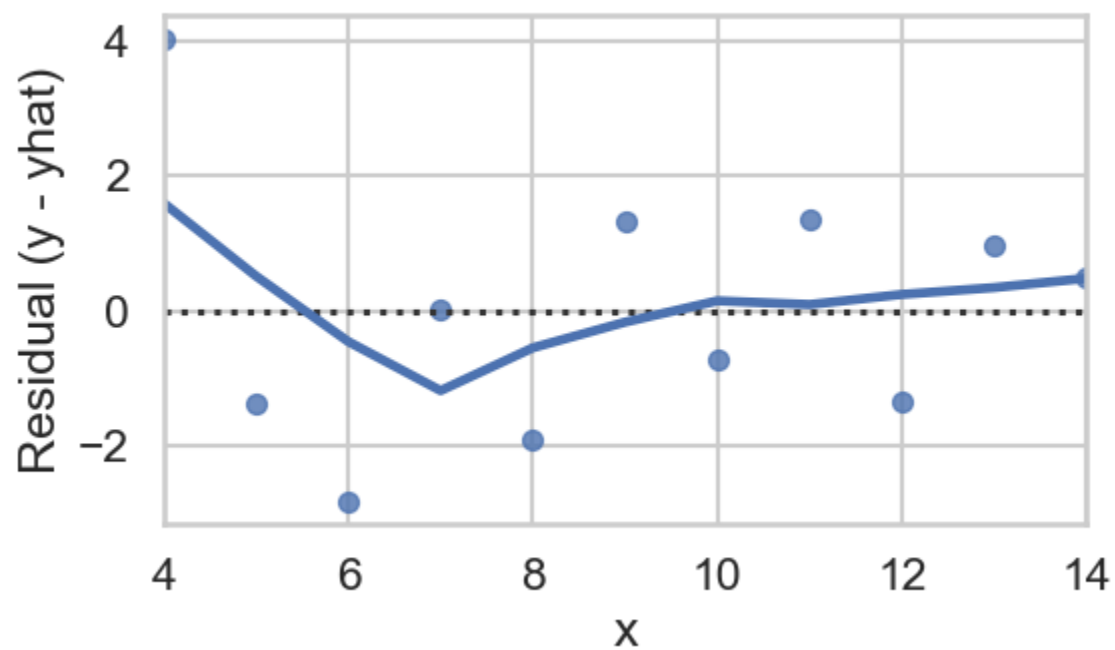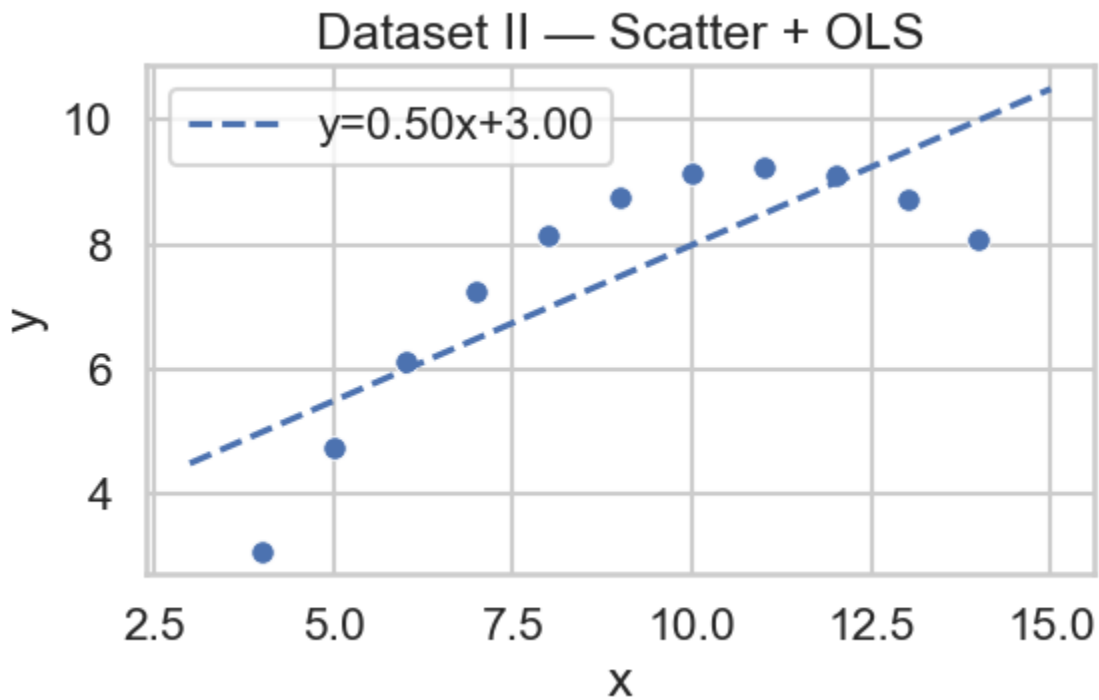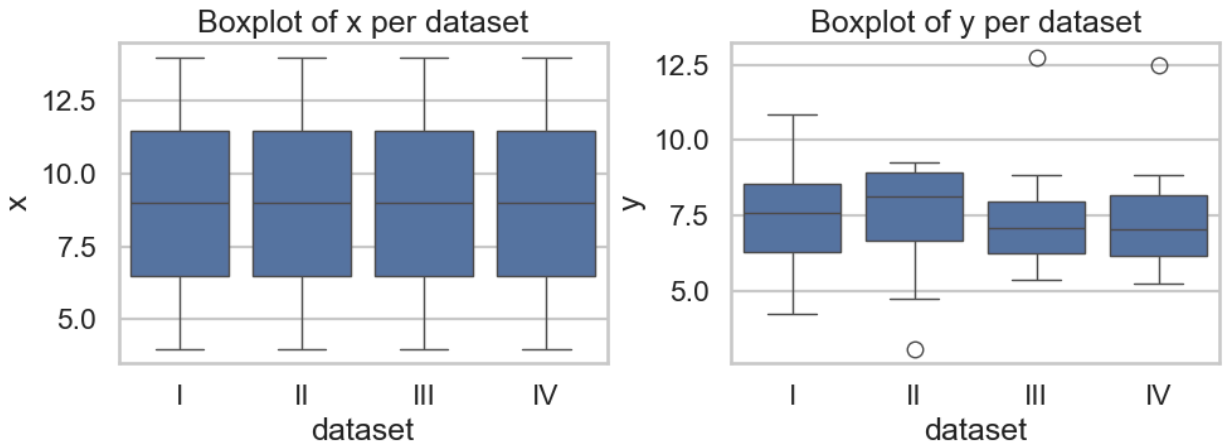
=== Anscombe's Quartet EDA Report ===

Title: Exploratory Data Analysis of Anscombe's Quartet
Author: Aayush
Date: 2025-10-09

Abstract:
This report analyzes Anscombe's quartet using summary statistics and visualizations. Although all datasets share identical statistics, their scatterplots reveal different patterns, emphasizing the importance of visual EDA.

Introduction:

Anscombe's quartet consists of four datasets (I–IV) with identical mean, variance, correlation, and regression coefficients. Purpose: Show why visualization is necessary alongside statistics.

Methods:
- Compute mean, variance, standard deviation, covariance, correlation, regression slope/intercept, R^2
- Scatter plots + regression lines
- Residual plots
- Boxplots / violin plots
- Interactive plots with Plotly and Altair

Summary statistics per dataset:

	n	mean_x	mean_y	var_x	var_y	std_x	std_y	cov_xy	r_xy	\
dataset										
I	11.0	9.0	7.5009	11.0	4.1273	3.3166	2.0316	5.501	0.8164	
II	11.0	9.0	7.5009	11.0	4.1276	3.3166	2.0317	5.500	0.8162	
III	11.0	9.0	7.5000	11.0	4.1226	3.3166	2.0304	5.497	0.8163	
IV	11.0	9.0	7.5009	11.0	4.1232	3.3166	2.0306	-2.115	-0.3140	

	slope	intercept	r2
dataset			
I	0.5001	3.0001	0.6665
II	0.5000	3.0009	0.6662
III	0.4997	3.0025	0.6663
IV	-0.1923	9.2314	0.0986

Interpretation:
Although all datasets share nearly identical summary statistics, their scatterplots reveal very different patterns:
- Dataset I: Linear trend
- Dataset II: Non-linear
- Dataset III: Vertical outlier
- Dataset IV: Single leverage point affecting regression
This demonstrates why visualization is essential in exploratory data analysis (EDA).