MONASH University

MONASH BUSINESS SCHOOL

Human vs. Computer

Shuofan Zhang

5/28/2018

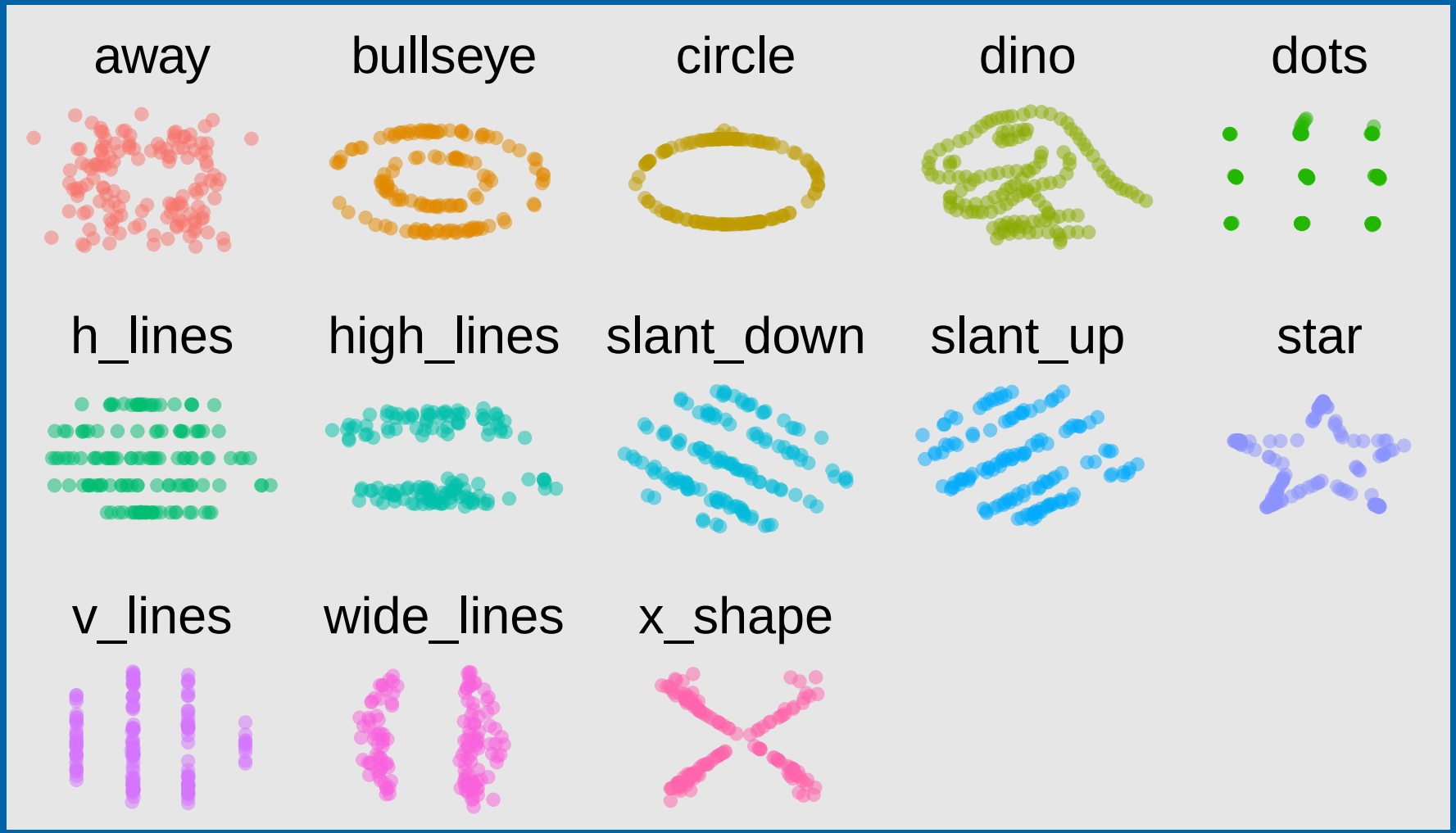
Reminder of the first presentation

Teach the computer to read residual plots

A major component used to diagnose model fits is a plot of the residuals. Residual plots are used to assess:

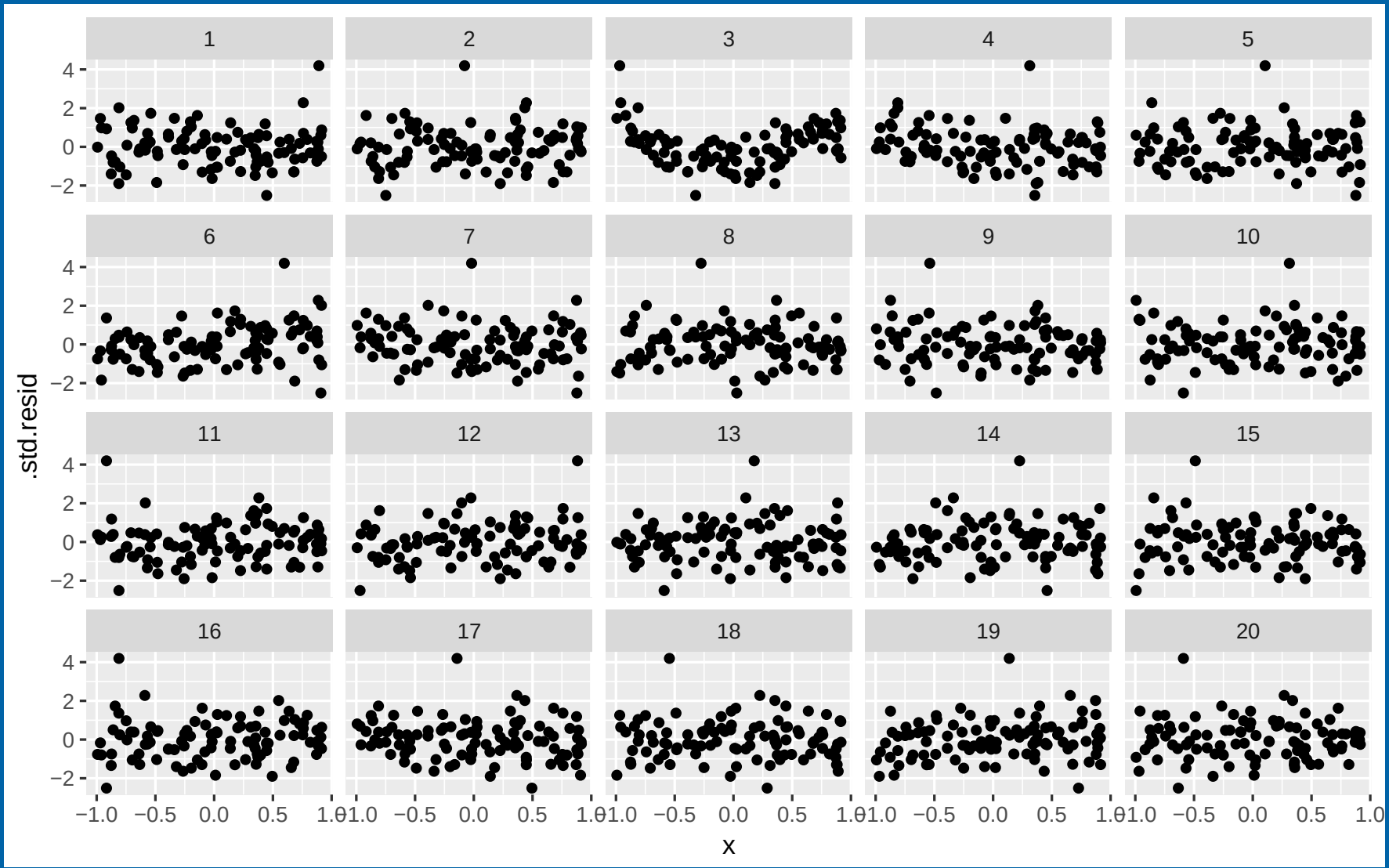
- Gauss-Markov assumption
- Heteroskedasticity
- Clumps of outliers

Why plots?



$E(x) = 54.3, E(y) = 47.8, sd(x) = 16.8, sd(y) = 26.9, r = -0.06$

Visual inference



Visual inference

- Plot of data is a test statistic
- Type of plot determines null hypothesis, e.g. residuals vs fitted scatterplot would imply H_o : no relationship, vs H_a : some relationship
- Human visual system evaluates lineup of data plot in field of null plots
- If data plot is “identified” as different from null, H_o is rejected
- Combining results from multiple observers enables p-value calculation

Deep learning

- Computer vision has advanced substantially
- Computer vision underlying self-driving cars, robotics
- Computer vision is being build on deep learning models
-
- If we can train a computer to read residual plots we can have it process a lot more data, than a human can manage.

Aside: Volvo admits its self-driving cars are confused by kangaroos

Volvo admits its self-driving cars are confused by kangaroos

Volvo’s self-driving car is unable to detect kangaroos because hopping confounds its systems, the Swedish carmaker says.

Aside: Computers can't tell difference between blueberry muffins and chihuahuas

Figure 1: Computers can't tell difference between blueberry muffins and chihuahuas

Experiment

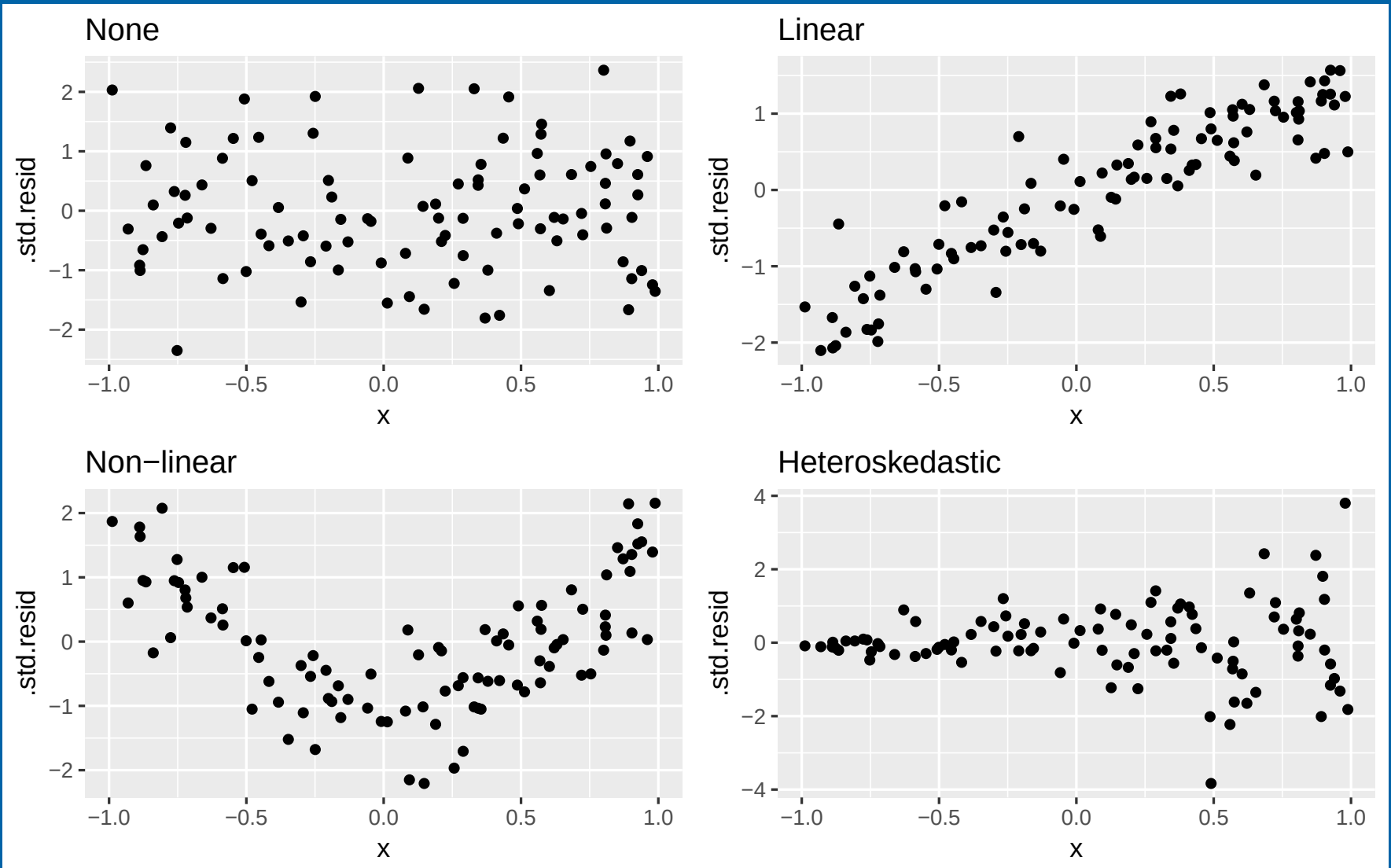
1. Simulate data from the different models
2. Fit a linear model to the data, extract standardized residuals and fitted values
3. Save residual plots as fixed-sized images
4. Train a deep learning classifier to recognise the departures from assumptions
5. Test the model’s performance on new data and compute the accuracy

Data simulation

These factors are being controlled in the data simulation

- Type of relationship: none, linear, nonlinear or heteroskedasticity
- Explanatory variables: $X \sim N(0, 3)$ and intercept $\beta_0 = 0$
- Sample size: randomly generated between 20-1500
- Image size: fixed 150x150

Type of relationship



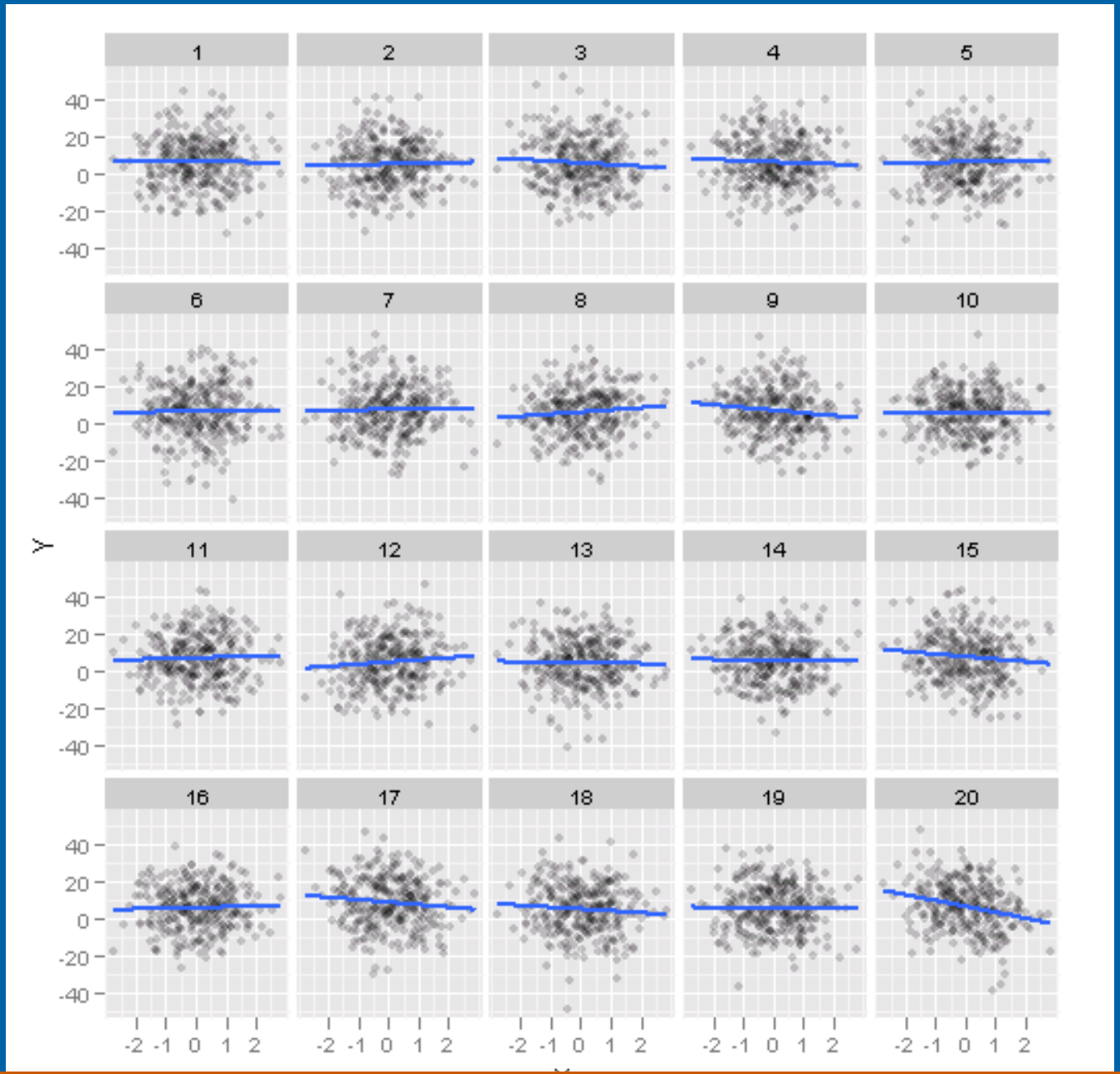
Simchoni's analysis

- Simulate data from linear relationship with $\rho = -0.9$ *to* $\rho = 0.9$
- Separate into two groups: significant/insignificant by t-test
- Train deep learning model with these two groups
- Test the model on lineup
- Successful in detecing linear relationship but fail in non-linear
- Simchoni's blog

Comparison with human subject experiments

- Majumder et al (2013) conducted a large study to compare the performance of the lineup protocol, assessed by human evaluators, in comaprison to the classical test
- Experiment 2 examined $H_o : \beta_k = 0$ vs $H_a : \beta_k \neq 0$ assessing the importance of including variable k in the linear model, conducted with a t-test, and also lineup protocol
- 70 lineups of size 20 plots
- 351 evaluations by human subjects
-
- Trained deep learning model will be used to classify plots from this study. Accuracy will be compared with results by human subjects.

Example lineup from experiment 2



Timeline

Date	Component
Apr 27	Deep learning model trained
May 4	Classification of new residual plots with model and results summarised
May 18	Comparison with Turk studies
May 24	Refinements made, final summaries written
May 31	Thesis finalised

Materials

- The thesis, code and data is available on the github repository https://github.com/shuofan18/ETF5550
- Software used to conduct this research is R, Tensorflow, keras, tidyverse