

Association for Information Systems

## AIS Electronic Library (AISeL)

---

AMCIS 2022 TREOs

TREO Papers

---

8-10-2022

### Applying Python Swarm Plot to Gain Insights into Business Data

Fang Chen

University of Manitoba, Fang.Chen@ad.umanitoba.ca

Follow this and additional works at: [https://aisel.aisnet.org/treos\\_amcis2022](https://aisel.aisnet.org/treos_amcis2022)

---

#### Recommended Citation

Chen, Fang, "Applying Python Swarm Plot to Gain Insights into Business Data" (2022). *AMCIS 2022 TREOs*. 49.

[https://aisel.aisnet.org/treos\\_amcis2022/49](https://aisel.aisnet.org/treos_amcis2022/49)

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2022 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Applying Python Swarm Plot to Gain Insights into Business Data

TREO Talk Paper

Fang Chen

University of Manitoba, Canada  
fang.chen@umanitoba.ca

## Abstract

Business, large or small, create visuals to facilitate decision making about operation and management. The most commonly used visuals can be grouped into five types based on the purposes: comparison, time trend, part-whole relationship, correlation, and distribution. The classical charts for comparison are column and bar charts. Line charts are the best to illustrate the time trend. Pie charts, donut charts, tree map charts, and waterfall charts are used to illustrate part-whole relationship. Scatter plots are for correlation. Histogram, box plots, and violin plots are used to show distribution.

Swarm plot is a new type of distribution chart that can be generated by Python and R. The swarm plot is more informative than other types of distribution charts, because it plots each of the data points on the chart, rather than just the summarized data. The following figure is a Python swarm plot for the survival rate in the Titanic tragic trip. The blue dots represent the survived passengers.

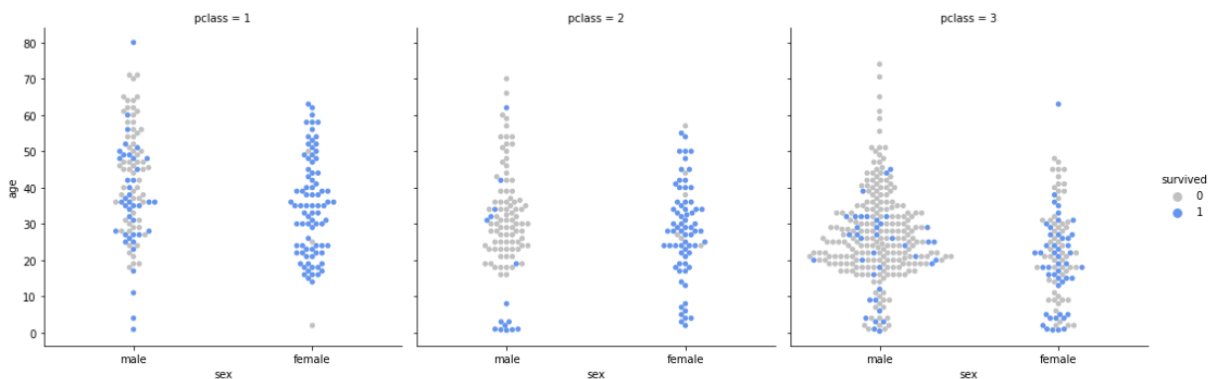


Figure. The Survival Rate of Titanic by Gender, Passenger Class (pclass) and Age

Swarm plots are very informative and versatile. However, the usage of this type of chart in business is not common. I believe it is worthwhile to share my experiences of how to use the swarm plots with other instructors who teach data visualization. To share my findings with conference participants, I made a PPT of swarm charts (with comparisons with other types of charts) for three datasets. I used swarm plots to tell the story of Titanic (passengers' gender, age, boarding towns, fare for different classes, and survival rate). The purpose is to illustrate how informative the swarm plots could be, then I created swarm plots and scatter plots with the diamond dataset (from Kaggle) to illustrate how color, clarity, carat and price are distributed and correlated in the dataset. As the conclusion, I used a retail store's dataset to gain insights into shoppers' shopping behavior by hours, day, and product categories. At the end of the PPT, I also include all codes that I wrote to generate the charts with Python Seaborn package. I will also share my observations about when it is good to use this chart, how to complement this type of chart with other types of chart, so that it is easier and more informative to tell a story with data. I will show the PPT slides to the conference participants.