**CIS 4170: Final Exam**

**Academic Integrity Statement:** By signing below, you agree that you will neither give nor receive unauthorized assistance on this exam.

# **Your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

# **Instructions:**

1. **Total Points: 175**
2. Download and save the data file for NYC Buildings Oil Consumption for year 2017.
3. On your machine, create a folder called 4170Exam. Please save all your data and python files in this folder.
4. Exam has two parts.
   1. Part I is conceptual and based on visualization principles.
   2. Part II is based on bokeh. After completing the exam, upload correctly named .py and .ipynb files on the blackboard using the assignment submission link by 3pm.
5. You have two hours to complete the exam. Good luck!

## **Part I: Viz Theory**

1. **[5 points]** Poppy was creating a thematic map with rainbow color scale for classifying her happiness data (% of people who are very happy) by counties. She should:
   1. Not use rainbow color scale as it her happiness data is quantitative and color does not provide ordering information
   2. Not use rainbow color scale as it her happiness data is categorical and color does not provide ordering information
   3. Use rainbow color scale because it rainbows symbolize happiness in people and doing so provides meaningful context as Tufte suggests
   4. Use shape instead of color visual variable to encode the % of happy people which is an ordinal variable
2. **[25 points]** Let’s say you want to create a visualization that shows sales of products that belong to different categories such as beverages, snacks, produce, and cereals. Each product category is likely to have 3-4 products.
   1. How many colors should one use and why?
   2. How should one go about selecting these colors to encode product sales across different categories?
3. **[30 points] From a visual perception and cognition perspective, identify and discuss at least 3 factors that can make visualization effective.**
   1. Factor 1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and Explain with an example why this factor can make visualization effective?
   2. Factor 2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and Explain with an example why this factor can make visualization effective?
   3. Factor 3: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and Explain with an example why this factor can make visualization effective?
4. **[10 points]** What is the difference between hover and linked brushing techniques?

## **Part II Bokeh: Using the NYC buildings Oil Consumption data, answer following questions**

Create a jupyter notebook called FinalExam\_FirstNameLastName.ipynb. In this notebook please create charts for Questions 5 and 6.

1. **[35 points]** In this Jupyter notebook, create a chart using that shows how many buildings will be retiring boilers in each year over a number of years. [Variables to use: BoilerRetirement\_dateEstimated, and BBL\_id]
   1. [20 points] Chart should have a dot (of any type: circle, square, etc.) that is clearly visible for each year and dots should be connected with a line. Hint: You will need to use two glyphs in the same chart (line and some type of ‘dot’) to create this.
      1. IMPORTANT: You do NOT need to change x\_axis\_type to datetime.
   2. [5 points] Style the charts as follows:
      1. Make sure dots and lines have at least following properties specified as needed: Size, color, nonselection\_alpha, nonselection\_color, and selection\_color
      2. X axis should be labeled as ‘Years’ and Y axis should be labeled as ‘'Number of buildings’.
      3. Chart title should be ‘Summarizing NYC Buildings Boiler Data’
      4. Format X and Y axis tick marks as needed including label orientation
   3. [10 points] Add plot tools as follows:
      1. Add hover tooltips to show
         1. the year and
         2. the total number of buildings
      2. Also include only following plot tools: lasso select, box select, pan, reset, and tap
2. **[50 points]** In this same Jupyter file, create two charts as follows [Variables to use: BuildingArea, TotalConsumption\_HighEstimateMMBTUs, TotalConsumption\_LowEstimateMMBTUs]:
   1. [20 points] First chart should show by building area, what is estimated high BTU consumption, and the second chart should show by building area, what is estimated low BTU consumption.
   2. [5 points] Style the charts as follows:
      1. Make sure your chosen glyph has at least following properties specified as needed: Size, color, nonselection\_alpha, nonselection\_color, and selection\_color
      2. X axis should be labeled as ‘Building Area in SF’ and Y axis should be labeled as ‘Fuel consumption in BTUs’.
      3. Chart titles should be ‘Examining Estimated High BTU Consumption by Building Area’ and ‘Examining Estimated Low BTU Consumption by Building Area’, respectively.
      4. Format X and Y axis tick marks as needed including label orientation
   3. [10 points] Add plot tools as follows:
      1. Add hover tooltips to show
         1. Facility Address
         2. Compliance with Greener Buildings Laws
         3. Primary Fuel used in the boilers
         4. Type of Building
         5. Neighborhood [Use variable: NTA]
      2. Also include only following plot tools: lasso select, box select, box zoom, wheel zoom, pan, reset, and tap
   4. [20 points] Create these two charts such that they
      1. share their both x and y range to allow for linked panning
      2. allow for linked brushing, and
      3. appear next to each other (and NOT one below another).

Create a .py file with following name to complete Question # 7 (and extra credit question # 8 if you wish to solve it): FinalExam\_FirstnameLastName.py

1. **[20 points]** In the chart you created in the Q #4 (two-glyph chart), add a SELECT widget that will allow a user to select from a dropdown [Variables to use: BoilerRetirement\_dateEstimated, ComplianceDate, BoilerInstallationDate]. In short this will allow users to view:
   1. how many buildings will be retiring boilers in each year over a number of years
   2. how many buildings will be compliant in each year over a number of years
   3. how many buildings have installed boilers each year over a number of years
2. **[30 points -- Extra credit question]:** In the same .py file write your code to create a chart that allows users to see total number of buildings within each buildingType for each of the fuel category [Fuel category to be selected from a dropdown]. The chart and its widget should appear below the chart and widget you created in question # 6.
   1. [20 points] Create the chart using appropriate glyph such that buildingType is sorted by the number of buildings in the descending order.
      1. Add a SELECT widget that will allow a user to select from a dropdown a fuel type
   2. [5 points] Style the chart as follows:
      1. Make sure your chosen glyph has at least following properties specified as needed: color, nonselection\_alpha, nonselection\_color, and selection\_color
      2. X axis should be labeled as ‘Types of Buildings in NYC’ and Y axis should be labeled as ‘Total number of Buildings’.
      3. Chart title should be ‘Which buildings use what fuel?’
      4. Format X and Y axis tick marks as needed including label orientation
   3. [5 points] Add plot tools as follows:
      1. Add hover tooltips to show
         1. Type of Building
         2. Fuel used
         3. Total number of buildings
      2. Also include only following plot tools: pan, reset, and tap.