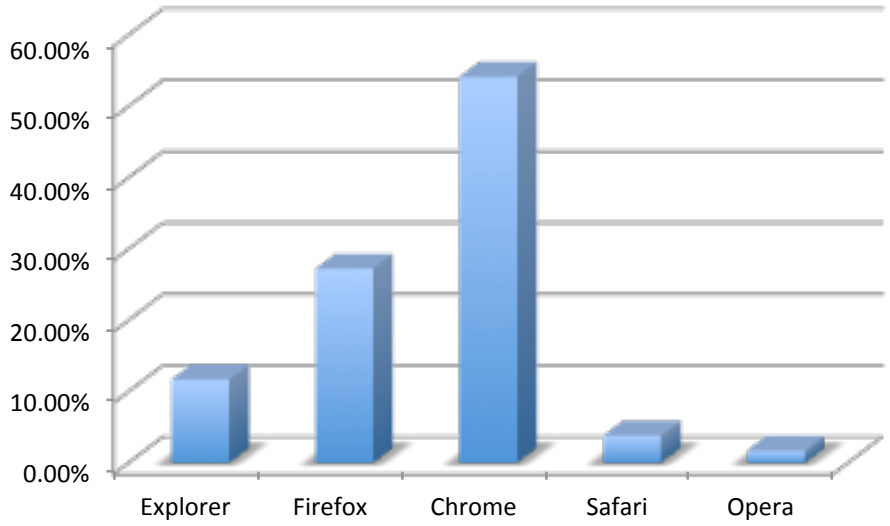
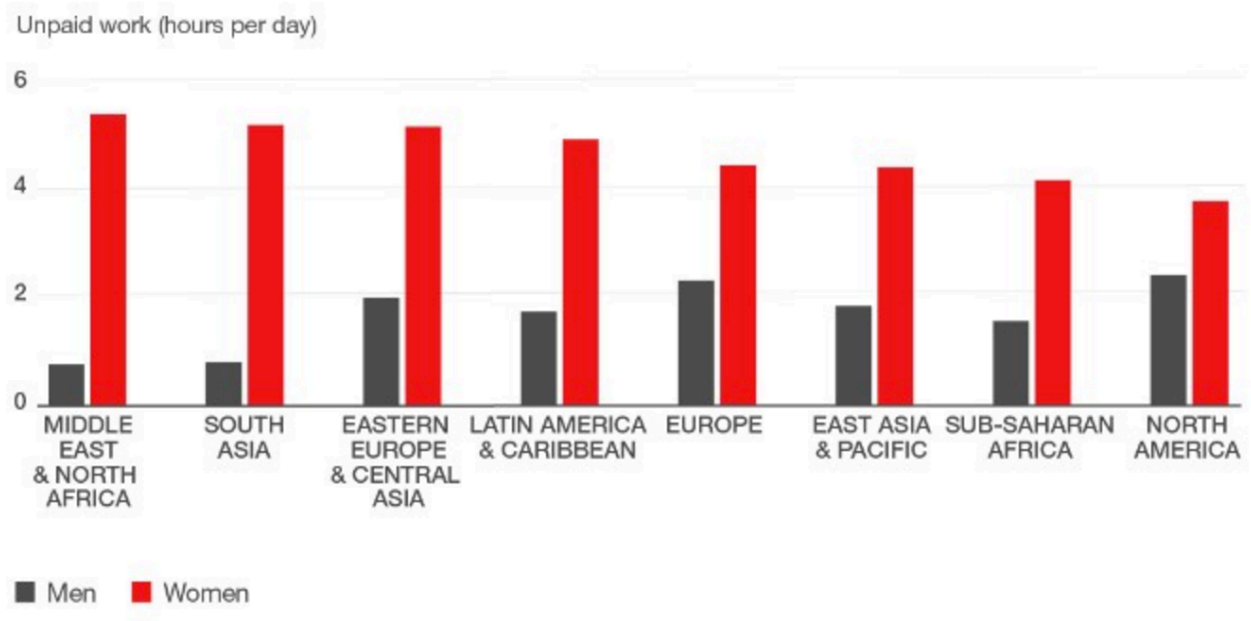
Assignment 2

Statistics 4420/8426
Exploratory Data Visualization and Quantification
Due on: Feb 22, 2017

1. Based on the Tufte's rule or Cleveland's principles and our discussions in the class, identify at least eight **distinct** problems of the plot below. Explain each of the problems you identified in terms of the specific rule it is violating.

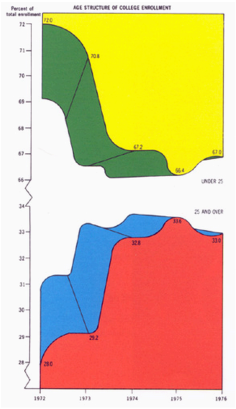


2. Based on the 13 criteria discussed in class, evaluate the following plot in terms of each criterion.



Picture Source: http://www.cnn.com/2016/02/22/opinions/bill-gates-melinda-gates-superpower-wish/index.html

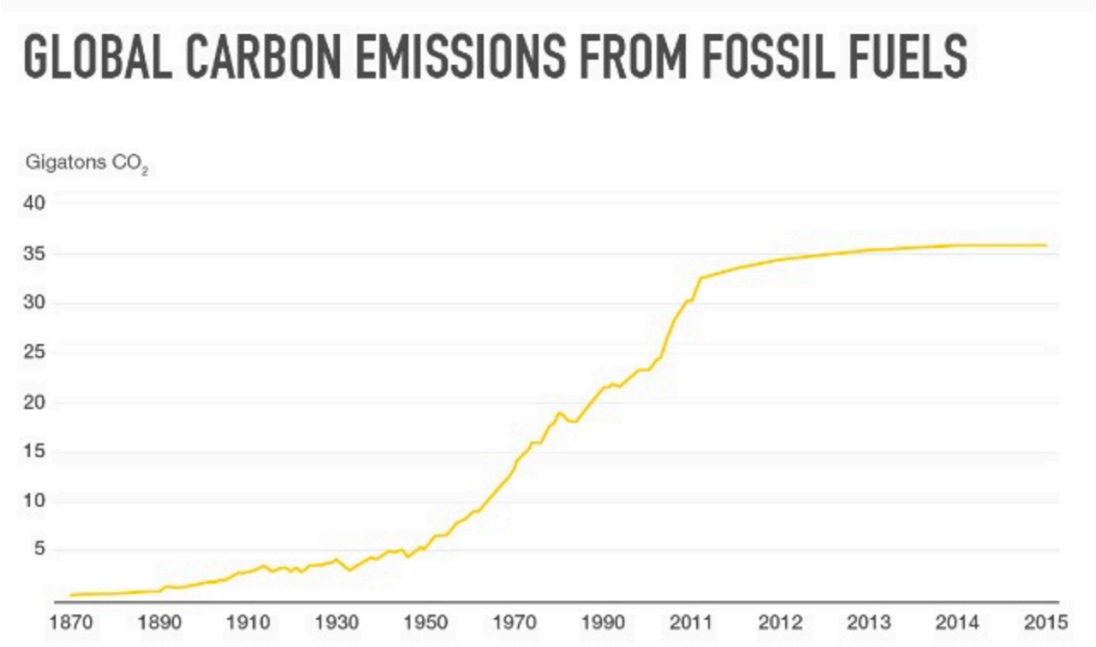
3. Write down the Tufte's rule for an effective graphical display of data. Which rules are violated in the following picture?



4. Identify at least five problems for the following display of data about fish population in a certain lake.

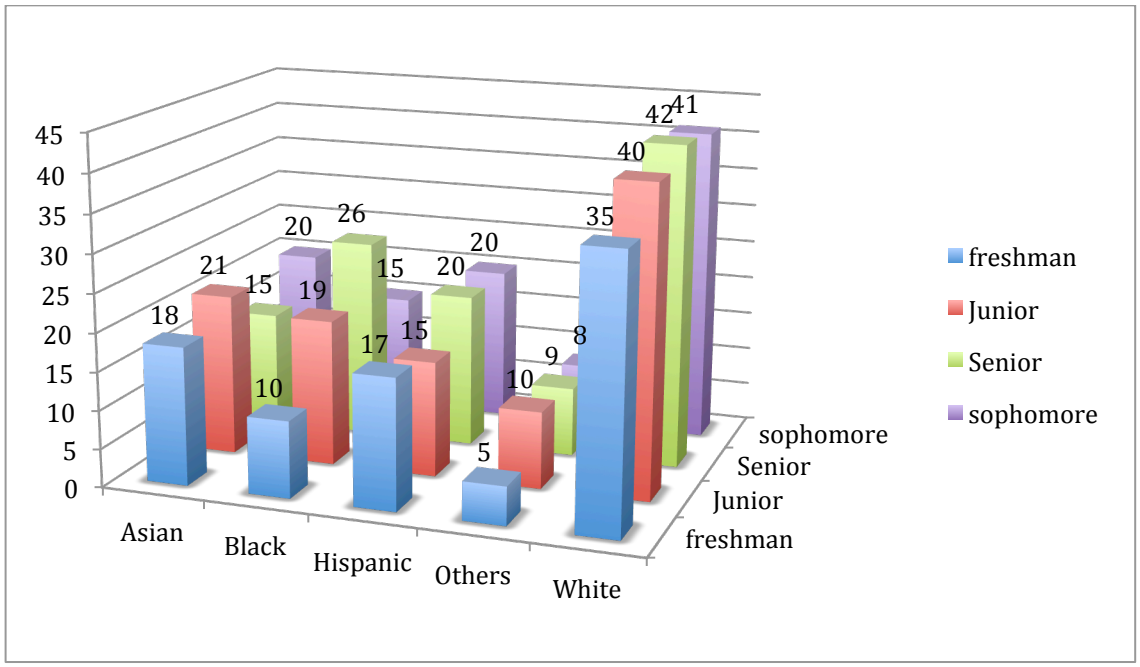
	Walleye	Cat fish	Crappie	Bluegill	Carp	% of fishes caught each year	% fishes die each year
1978	239,587	56,877	3,984,745	9,856	597	28.456	25.24%
1979	432,576	43,597	4,459,847	2,356	27856	39.340	22.34%
1980	322,873	45,698	3,348,756	4,256	2395	29.352	27.67%
1981	674,598	22,454	39,834,875	2,856	384	35.465	27.45%
1982	2,358,745	34,876	9,834,274	2,398	43937	24.782	28.67%
1983	753,684	29,384	6587,324	2,396	6257	56.351	29.76%

5. CNN just published the following picture on Feb 22, 2016. Compute the lie factor in the picture. Discuss what impact we observed due to the lie factor.



Picture Source: http://www.cnn.com/2016/02/22/opinions/bill-gates-melinda-gates-superpower-wish/index.html

6. Answer the following questions based on the plot below. The plot shows the enrollment based on class and race.



a) We can see the data in the plot. Create a data frame and turn in your data.
b) Do you think the plot is the best display of the data? If not, recreate one of the best plots of the data using any software and turn in your plot.
c) Comment on the data ink ratio of the plot
d) Generate a plot showing average enrollments by race. Do the same for class.
e) Do you think that race is significant to explain the variations in enrollment? Why?
f) Do you think that class is significant to explain the variations in enrollment? Why?