누구나 캐글에 입문할 수 있다. (feat. 지방대/비전공/인문학도)

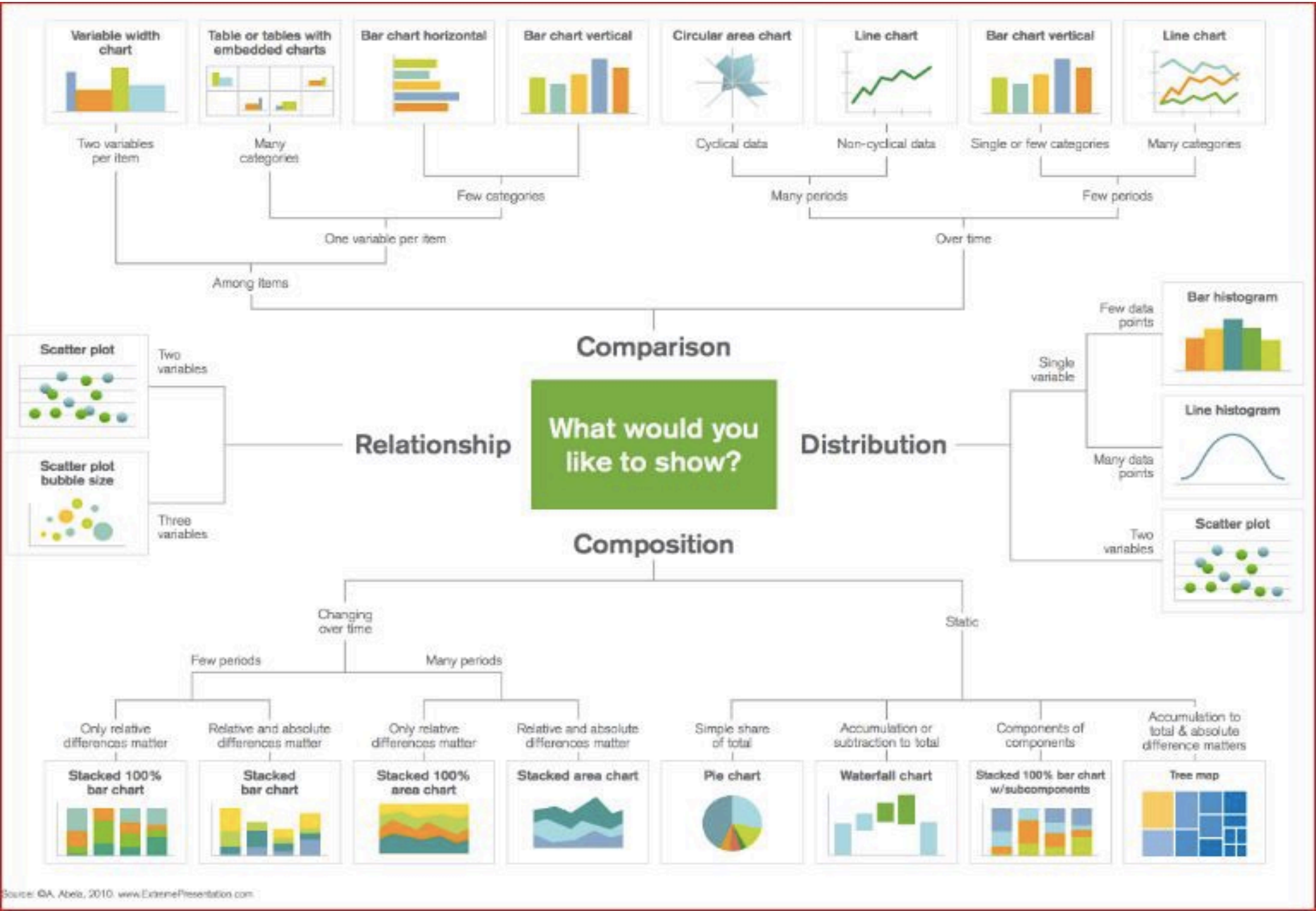
- 이상치의 개념

j2hoon85@gmail.com

2-11 Graph with Outliers

이상치 판별 그래프

j2hoon85@gmail.com



Source: @A. Abela, 2010. www.ExtremePresentation.com

핵심은 평균과 실제 관측값과의 차이다!

	Population	Sample
# of subjects	N	n
Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

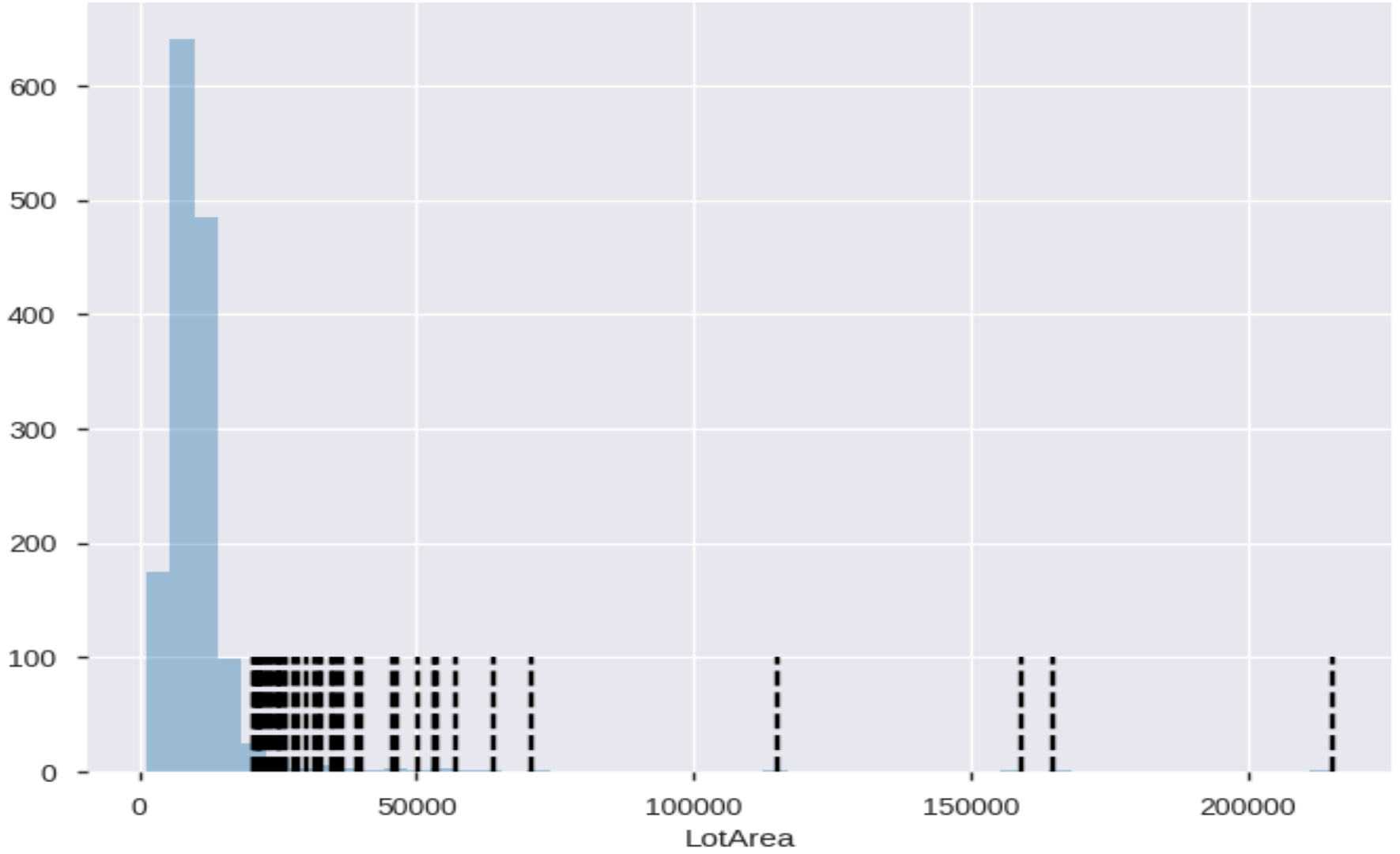
Note: S^2 is the formula for unbiased sample variance, since we're dividing by $n-1$.

Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

Note: Finding S by taking $\sqrt{S^2}$ reintroduces bias.

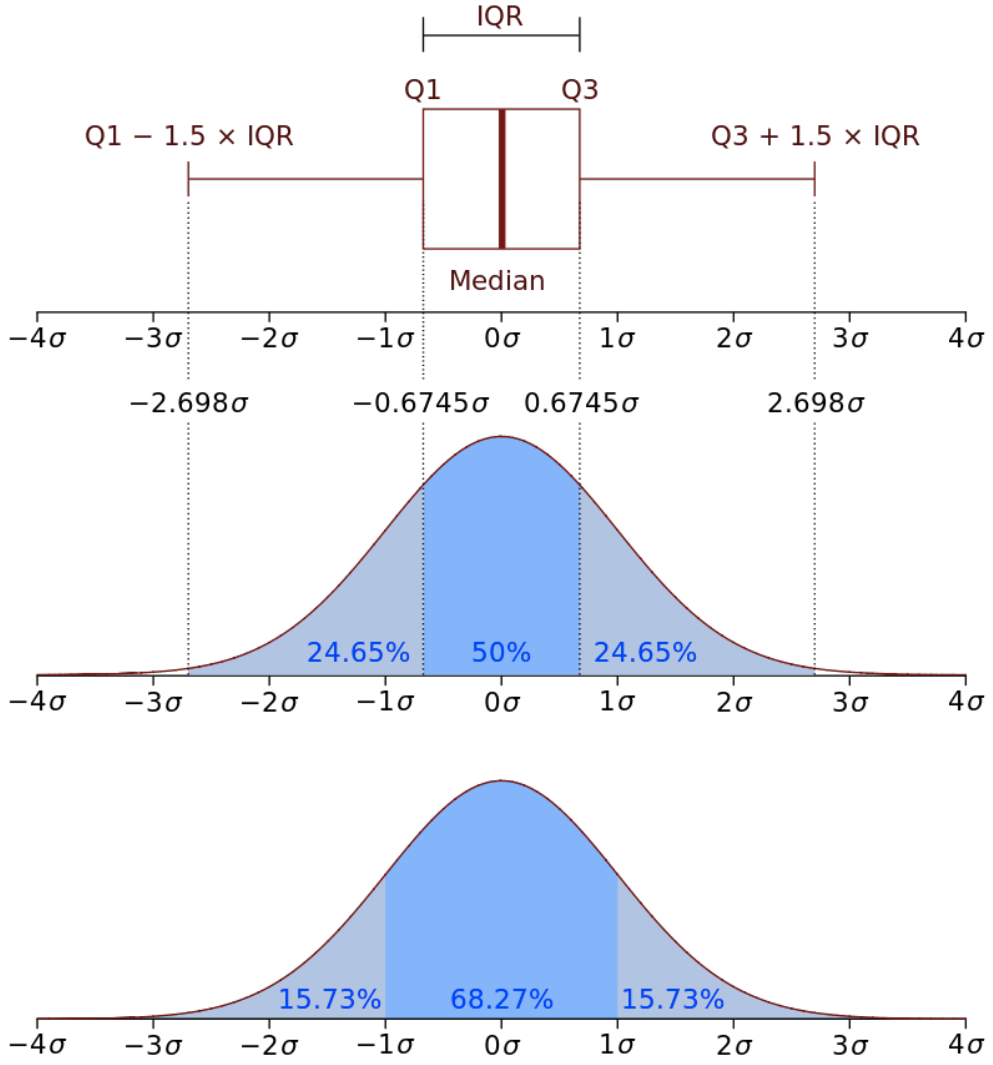
j2hoon85@gmail.com

표준편차를 활용한 그래프



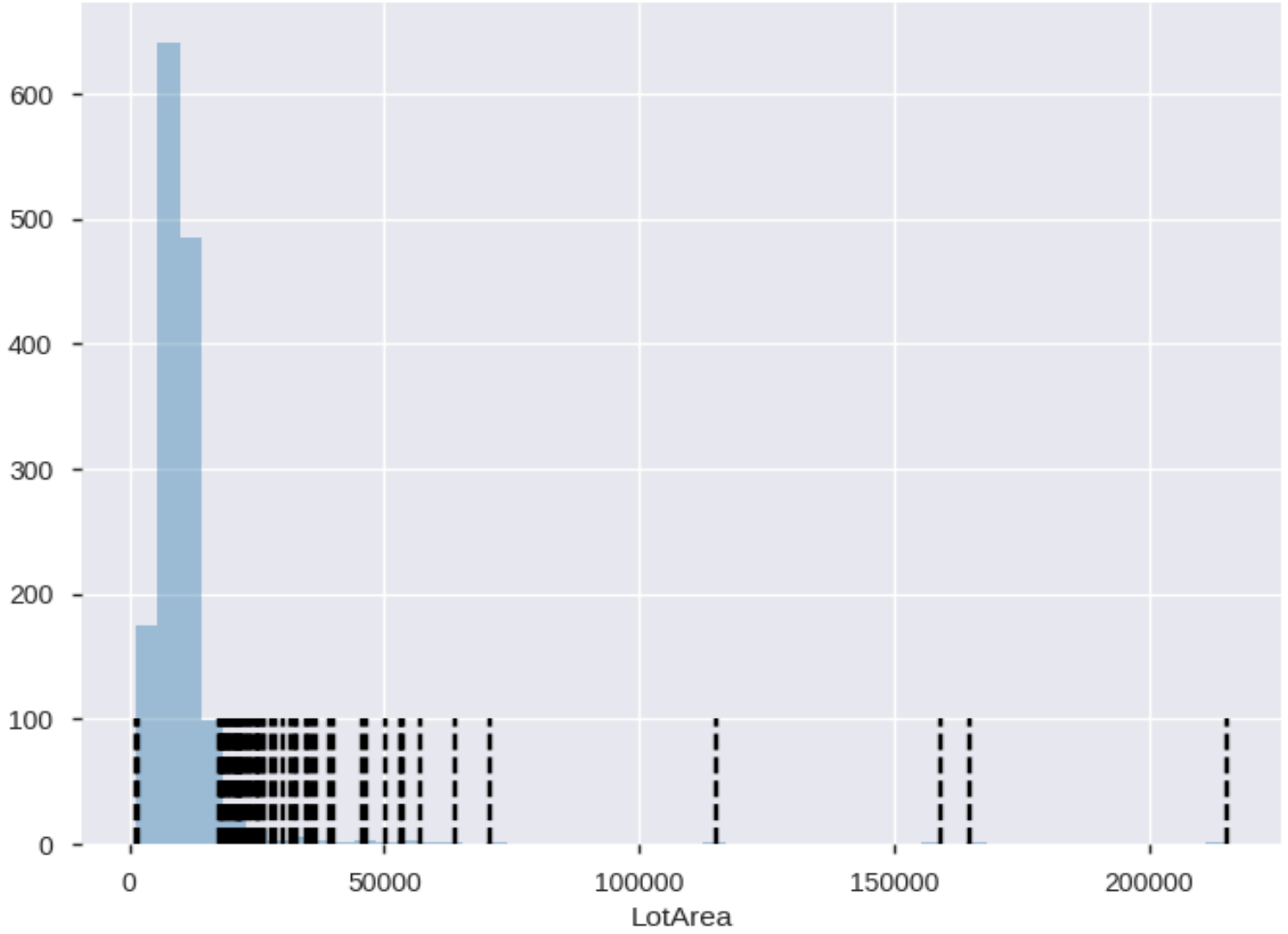
j2hoon85@gmail.com

사분위를 활용한 그래프



j2hoon85@gmail.com

사분위를 활용한 그래프



j2hoon85@gmail.com

Happy To Code

j2hoon85@gmail.com