W8和TF题（考前再看），W3代码，W5， W9-10，W12-13，W14

W2：

1，Data Science Process (6个)，get data（3个）

2，atomic data (3个) ，atomic data组合（3个），data Storage（3个），Data Format（3个），Types of Data（1-2，1个）

3，population 和 sample，selection bias 和 Volunteer/nonresponse bias

4，mean 和median，Skewness，Range，Variance 和 Standard deviation

W3：

5，***DATA CLEANING benefit***（5个），DATA CLEANING CYCLE（8个）

W5

6, ***PRINCIPLES OF VISUALIZATIONS***（4个），TYPES OF VISUALIZATIONS（4个）和***主要的charts***（2，2，2，2）

W9-10：

7，5种feature，FEATURE SELECTION（3个）

8，***How to draw the first and second principal component in PCA method***

W14：

9，***Distance Measures***（2个）

余弦距离在余弦相似度前面加 arc（即cos-1）

单词：

Exploration, Tabular, Temporal, Categorical, normalize, series, variation, Comparison, bubble, histogram, quartiles, spatial, binarization, Embedded, aligned, perpendicular, continuity, observation, predictor, Multicollinearity, Dimensionality, covariance, eigen-vector, Gradient Descent, Ensemble, Bootstrap aggregating, out-of-bag error, decision stumps, Collaborative filtering, similarity, interpretability

