

THE HEWLETT FOUNDATION
Automated Short Answer Scoring
Data Visualization Contest

A Story In Pictures

© 2012 Nitin Madnani

Introduction

The William and Flora Hewlett Foundation recently organized a public competition called the Automated Student Assessment Prize (ASAP). The goal of the competition was to “*solve the problem of the high cost and the slow turnaround of hand scoring thousands of written responses in standardized tests. As a result many schools exclude written responses in favor of multiple-choice questions, which are less able to assess students’ critical reasoning and writing skills.*” The competition was organized in two phases. The first phase asked participants to build systems that can automatically score essays-length answers (150-550 words) whereas the second phase was focused on short answers (approx. 50 words). This second phase provided human graded answers written by students for 10 different short answer questions as training data to the participants. The 10 questions (or *tasks*) ranged from subjects such as Biology to Language Arts. More details can be found on the official Kaggle webpage for the contest [1].

In addition to the primary machine learning competition for short answer scoring, a secondary contest was also organized as part of the second phase where the goal was to come up with interesting ways to visualize the data provided to the participants [2]. This document represents an entry for this data visualization contest. It was created using a combination of Adobe InDesign and R.

Description

The goal of this entry was to visualize different aspects of the data using a multitude of visualization techniques and present it as three different chapters of a story.

The first chapter visualizes the statistics about each task such as average length of responses per score point, score distribution and the human-human agreement. This clearly shows that tasks 5 and 6 have an overwhelming number of zeros and that length is likely to be a very good baseline predictor of scores for almost every task. The second chapter contrasts the words used in the highest scoring responses with the words used in the lowest scoring responses, for each of the 10 tasks. Finally, the third chapter illustrates the extent of misspelling (or transcription) errors in the data for the most frequently misspelled words. It’s hard for a spellchecker out there.

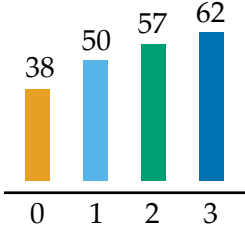
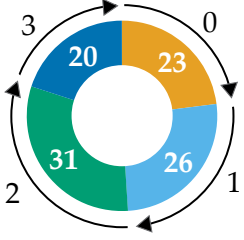
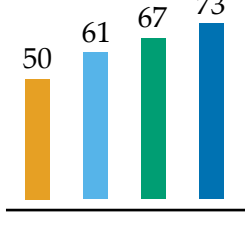
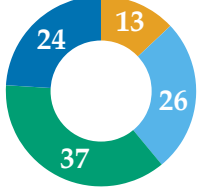
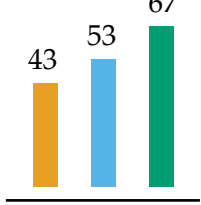
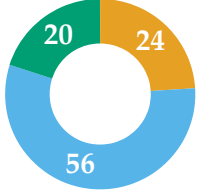
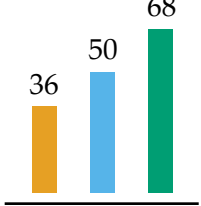
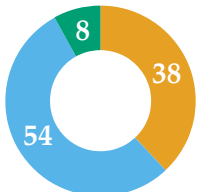
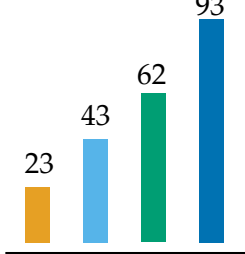
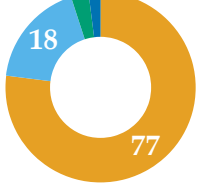
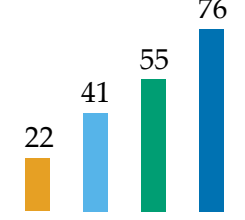
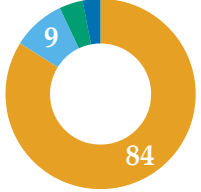
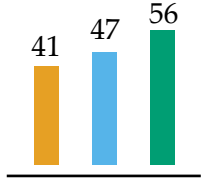
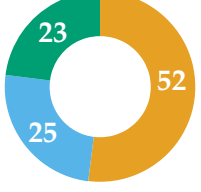
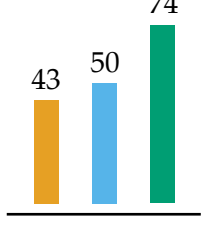
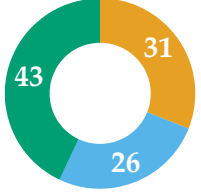
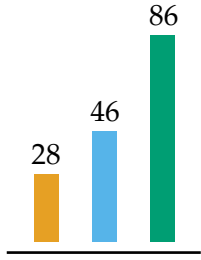
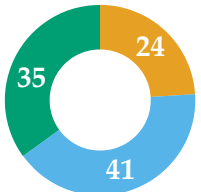
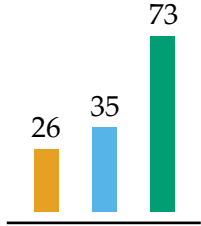
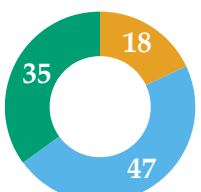
References

[1] The Hewlett Foundation: Short Answer Scoring (<https://www.kaggle.com/c/asap-sas>)

[2] Visualization Contest (<https://www.kaggle.com/c/asap-sas/prospector>)

CHAPTER I

“A swampy slurry of bits of hard data and buckets of mushy manipulation”
 (Laura Penny)

TASK	AVERAGE WORDS PER SCORE POINT	SCORE DISTRIBUTION (%)	HUMAN AGREEMENT (K)
1			.86
2			.80
3			.65
4			.68
5			.91
6			.89
7			.93
8			.75
9			.71
10			.81

CHAPTER III

“I don't give a damn for a man that can spell a word only one way.”
(Mark Twain)

temperature

tempature
temperture
tempurature
tempeture
temprature
temperatu
temerature
temputure
temperatur
tempatures
teperature
tempreture
temputature
tempriture
tempertaure
tempertures
temparature
tempreature
tepature
temurature
tempurture
temerture
temerature
temperater
temperatre
tempture
temperaute
tempenture
tempeature
tempatue
tempurtue
tempatur
tempuature
temperatuire
temapture
tepmature
tempatuer
tempayure
tempetures
temporter
tempurate
ttemperature
temerature
tempurat
temrature
tmeperature
tempereture
temperatrure
temperature

↑
Misspelling Frequency

different

diffrent
differnt
differant
diferent
diffent
differnet
differet
diffferent
dfferent
diiferent
difrent
differat
differt
diferrent
diffrenet
driffent
defferent
distrant
diffrant
differend
diffirent
deffent
diffenert
diffient
dffent
diffren

nucleus

nucleous
nuclues
nucleas
nucleaus
nucles
nuclus
nucleos
nucleos
nuleus
nucleus
nuclueus
nuclius
nuclesous
necleous
neucleus
nucleus
nucluelus
nucleolous
nuclis
nucleos
nuclieus
nueclus
necleus
nucleous
mucleus
nucleuas
nucleuse
necluus

experiment

expiement
expiement
experement
expirament
experament
expariment
experient
experment
exeperiment
expierement
expirment
exsperiment
expriment
expierment
eperiment
expeirement
expariement
experamint
exerinent
experiamt
exprement
expinment
expairement
exiperiment
expiement
expiement
exprement
expiement

information

infomation
infomation
imformation
information
infernation
informations
informatioon
informatio
iformation
informattion
informatin
informaton
informantion
information
informatiosn
nformation
infortmation
informatiom
informatioin
infarmation
informtion
niformation
informactin
unformation
informotins

article

artical
atricle
aticle
article
artiale
artivle
artichel
artcile
artilce
artictle
articol
article
artlicle
artclie
aritical
artcle
artcl
articke
artile

membrane

membrain
membrane
memebrane
mebrane
membran
memberane
membrans
mambrane
mebrain
membrane
membarin
membrane
membrian
memabrane
membrand
membran
mimbraint
menbrane
membrade
membrean
mambrain

environment

enviroment
enviorment
enviorment
environment
envorment
environment
enviornment
eviorment
envoment
envinroment
inviroment
envierment
envlromet
envirment
envirnoment
inviornment
envirioment
evorment
environmetn
envoriment
envoiment

species

speices
spieces
spieces
specias
specease
specites
speciese
spcies
speciece
speies
spiceies
speiees
speciecs
sipecies
speces
specia
speals
specis

Leonard

Leonard
Lenoard
Loenard
Lenoard
Leomard
Leonard
Leonared
Leanord
Leondard
Leoard
Leonard
Leonard
Leonard
Leonard
Leonard
Leonard
Leonard
Leonard
Lwonard

protein

protien
portein
protine
prton
proein
proteing
potein
proetin
prtein
proteine

Notes:

- In Chapter I, scores from the first rater are used to compute average number of words for each score point and the score distribution.
- In Chapter II, words from the prompt text are removed from the responses before generating the word clouds.