GCT561 Scientific concept and thinking

Name: Chaelin Kim

Student ID: 20205096

HW1: Probability
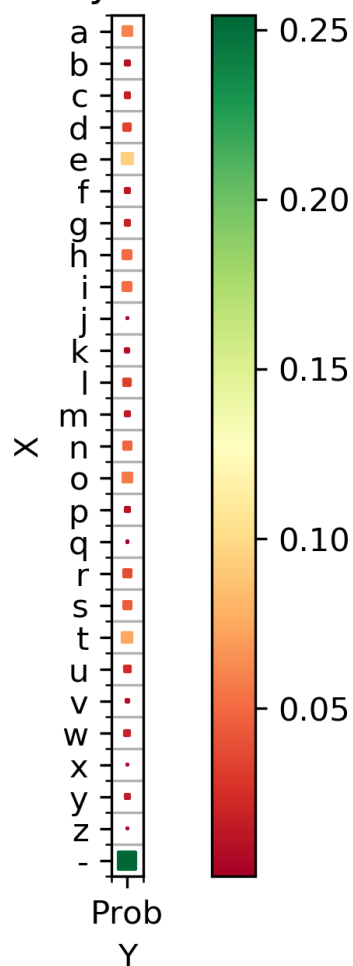
0. English text

- "Alice's Adventures in Wonderland by Lewis Carroll 1865" in nltk.gutenberg package
- Length of the text: 144,395

1. Find the probabilities of each character

- **Probability distribution P(x)**

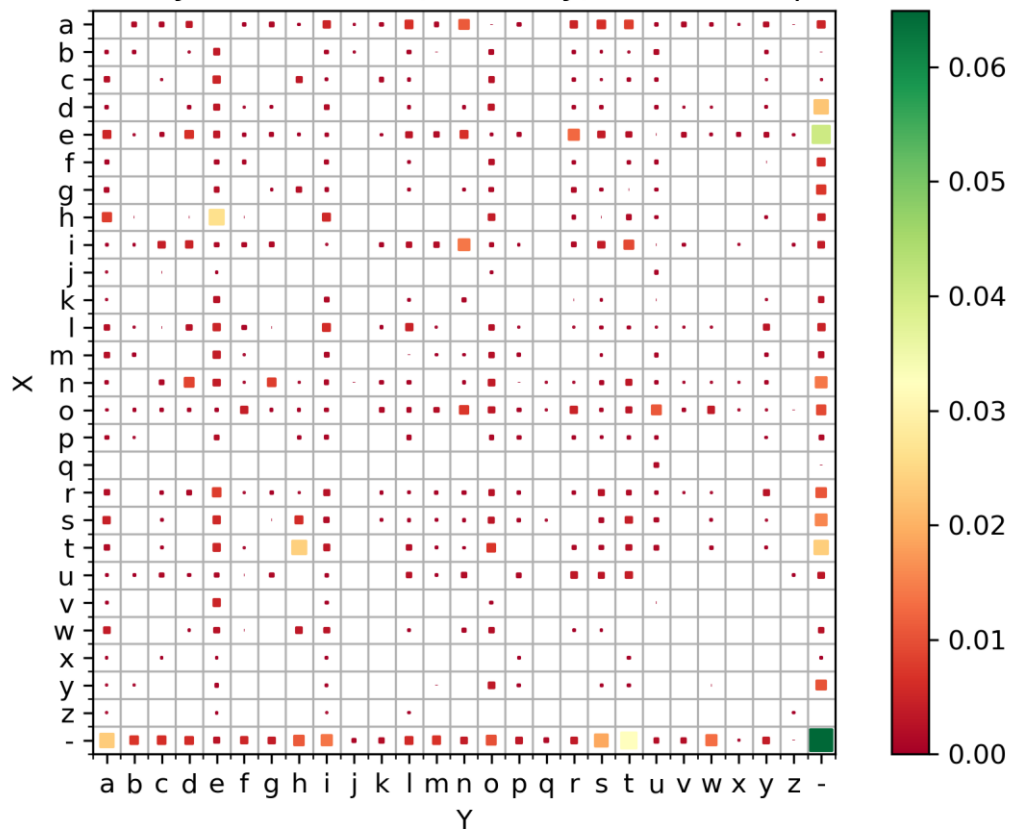|	Prob
a	0.06088
b	0.01022
c	0.01661
d	0.03414
e	0.09398
f	0.01385
g	0.01753
h	0.05105
i	0.05201
j	0.00101
k	0.00802
l	0.03264
m	0.01457
n	0.04856
o	0.05641
p	0.01055
q	0.00145
r	0.03765
s	0.04502
t	0.07400
u	0.02400
v	0.00586
w	0.01853
x	0.00102
y	0.01567
z	0.00054
-	0.25423

Probability distribution P(X=x)
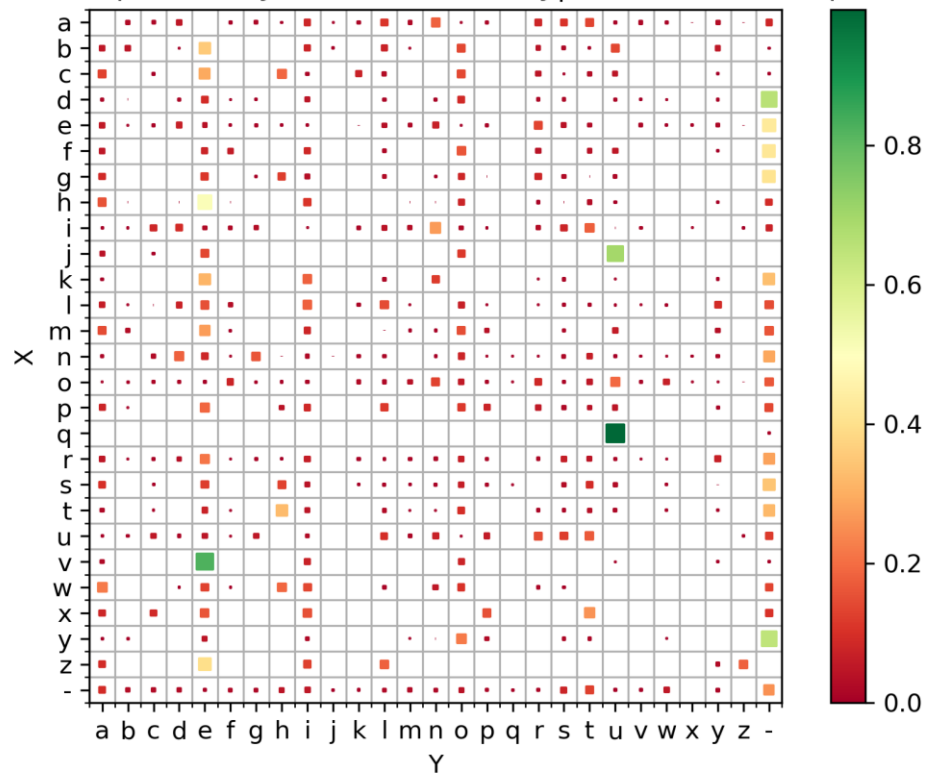for randomly selected letter

2. Find the probabilities of each bigram
   · **Probability distribution P(x,y)**

Probability distribution P(X=x, Y=y) for ordered pair XY

· Conditional probability distribution P(y|x)

Conditional probability distribution P(Y=y|X=x) for ordered pair XY

· Conditional probability distribution P(x|y)

Conditional probability distribution P(X=x|Y=y) for ordered pair XY