# Google's mission "To organize the world's information and make it universally accessible and useful,"

Google’s corporate vision statement has three variables, namely, **world’s information**, **accessibility**, and **one click**

**Vision:** Google's Android phones offers its users free access to Google Maps and provide navigation and communication. Google is also branching into other areas as well, as the famous Google cars that operate on their own. This is part of Google's vision of a future where cars drive themselves and people can participate in more productive activities during their commute.

The site Google+ is part of Google's vision for social media, which is integrated in the phones and all other aspects of life. Google Chrome browser is part of the company's vision of a more open market for Web browser add-ons. The innovative company has even started a new venture called Calico, which is focused on creating a healthier future for all mankind.

 The company fulfills the “*world’s information*” component of the vision statement by indexing webpages and maintaining databases of the resulting information. Google fulfills the “*accessibility*” component by offering its products to everyone around the world. The “*one click*” component of the corporate vision refers to easy access to information

Committed to significantly improving the lives of as many people as possible

# Introduction: Data Scientist, PhD University Graduate role

My name is Amin, I am a 5th year PhD student in statistics at Iowa State University with anticipated graduation date of summer 2022. I am available in the job market, and I am so excited to have this interview with google today because when I came across the vacancy ads, I said that’s the company I’d like to work for, and I could try my chance to get this position.

When I looked at the job description, I believe that’s something I could be beneficial for. Because my research is related to the multiple testing for high dimensional data, and my most recent internship experience in biostatistics, as well as my current consulting position along with my experience of collaborating with different departments such as sales and marketing make me think that I can be a good fit for the position and can add some values to the mission and vision of google along with my personal growth.

# Why google?

There are interesting problems and interesting people to work with, a corporate culture that encourages taking initiative, good pay and benefits, and the company tries to make a positive impact on the world.

# Consulting projects?

# Research: (motivation, method, advantages, result)

My research is on multiple testing of high-dimensional data when the variables are highly dependent. Usually for a treatment vs. control groups of data in an experiment, we’d like to see whether the population means/ratios are significantly different.

A good example for such data is gene expression data. Normally genes that are next to each other in gene pathways share more common information and that makes them correlated.

So, when considering a treatment effect, we may want to run multiple comparisons to identify the genes that are significantly different between the two groups. But, since we have somehow strong dependence among the genes, common methods such as two-sample t-test might lose power.

So, dependence is troublesome here, but what if we could exploit this dependence to come up with more powerful test. Our method is just using some simple linear models to take advantage of the dependence.

For each of the comparisons, we decided to run a multiple linear regression on the genes which are in a small distance from each other, where they are more probable to share the same information. Then proposed a test statistic which is based on the residuals of the regression model. Because the residuals usually have smaller standard deviation, we have a higher signal to noise ratio and can achieve more powerful tests.

This is a very simple idea which can be applied to the original data without any need to use any Bayesian or shrinkage methods to benefit from the dependence among the data. Another advantage is that, unlike some other procedures, we do not transform the original data to have higher signal to noise ratio.

Our simulation and empirical results have shown to be performing better than other available methods.

# Internships:

In summer 2021

Merck

1. Package development for Merck Oncology Survival Analysis called mksurv
2. Merck adaptive 2-in-one design package
3. Merck qualification package for the assessment of internal packages and regulatory purposes
4. Created 4 shiny apps for the packages I worked on.

In summer 2020

I worked as a biostatistics graduate intern in the Design methodology division at Merck

1. Running simulation to compare the asymptotic results for a fixed or group sequential design under non-proportional hazard assumptions
2. Package development for three different packages related to group sequential clinical design (gsdmvn, simitrial, and gsDesign2)
3. Conducted some research in design methodology for group sequential designs

In ClinChoice: Fall 2020 & Spring 2021

1. Developed unit testing for clinical trial packages to evaluate the functionality of the package
2. Simulation study for oncology clinical trials

End questions:

1. What kind of value are you looking for a person to add to your team?
2. What are potential challenges for a person in this position?
   1. What challenges did you face?
3. What would you describe a typical project for someone in this project?
4. Rather than some coding skills required to be successful in this position, what interpersonal skills do you think is necessary for such a position?
5. What is the next step?