**GENES INFLUENCING YOUR PERSONAL INCOME.**

By (name You can also include reg no.)

The Name of the Class(course)

Professor(tutor)

The Name of the School (University)

The City and State where it is located

The date

Introduction

Over the past, the gap between the rich and the poor has always been there. Many have tried to explain in many terms why the reason for this exists. Scientists all over the world are trying to ask themselves if there could be reasons that could lead to this that are related to parental or even the genes. In this research paper we are trying to answer the same question. But we are in a different way trying to look at the things that are associated with this. We are targeting the big question, “Are there genes influencing your personal income, and is it possible to identify these genes?”

Research that involved the study of the genes and the relations that it has to people making money is one of the researches that were done by the University of Edinburgh. In this research we find there is one David Hill who argues and he says “Your DNA will not print your money”. In the research they find the people that are in a certain area and they link the genetics in a statistical manner. Along with this there are also collaborators from the European and Hill also sifted through the data that was in the Biobank of about 286000 persons. This research by Davis Hill, focused on the polygenic score, a genetic calculation that predicts a person’s odds of reaching a certain outcome- of say developing diabetes or even earning six figures.

According to research by PsycInfo database records, they argue that genetics can also relate to economics in the manner that they suggest. In this relation with economics, it influences the income that someone can get. They say that measures of wealth such as income and assets are commonly considered to be objective measures of environmental circumstances, making direct contributions to life satisfaction. Using a nationwide sample of 719 twin pairs from the National Survey of Midlife Development in the United States, the authors first noted the relative independence of most perceptions about financial status from measures of actual wealth. They then demonstrated that perceived financial situation and control over life completely mediated the association between measures of actual wealth and life satisfaction. Finally, they showed that financial resources appeared to protect life satisfaction from environmental shocks. In addition, control appeared to act as a mechanism translating life circumstances into life satisfaction. (PsycInfo Database Record (c) 2021 APA, all rights reserved)

According to research carried by Jama Psychiatry, John N. and Richard D. argue that there is an influence that is generated by autism. Hence if the genes from the family can lead to such kinds of conditions, then there is a close relation to the way the social economy of someone. This is a direct influence in the income that someone may pose to have in the future. They used the twin study to draw conclusions of the fact that the social responsiveness scale.

In another study by the same article, we find Eric Schmitt and Kenneth S. elaborating on the influences that the genes and environment have on Alcohol, caffeine, cannabis, and nicotine from early Adolescence to middle Adulthood. While it can be true of the argument that they make, we have but little of these influences differ through development as well. In the research that they did they conclude that the support of aetiology for an individual differs in PSU in the initial and early patterns of the use of strong influence by the social and familial environment factors while later they are stronger influence. The factors that are mentioned are known to make people lag in their social economic status. This is thus the basis of the argument that the two have at the end of the study that they made.

Last, according to the nature reviews of neuroscience, Daniel, Mathar and Michael have reasoned that human brain development occurs within a socioeconomic context and childhood socioeconomic status (SES) influences neural development — particularly of the systems that subserve language and executive function. Research in humans and in animal models has implicated prenatal factors, parent–child interactions and cognitive stimulation in the home environment in the effects of SES on neural development. These findings provide a unique opportunity for understanding how environmental factors can lead to individual differences in brain development, and for improving the programs and policies that are designed to alleviate socioeconomic status-related disparities in mental health and academic achievement. These factors all come back to explain our main topic and thus in this research we are developing further what has been talked about by other scientists. As we look at the reason, we are going to make an analysis to make concrete some of the findings that are in this paper.

Twin Study

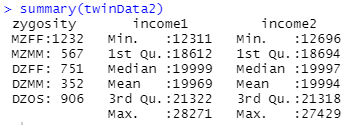
We define Twin studies as studies that are conducted on identical or fraternal twins. They aim to reveal the importance of environmental and genetic influences for traits, phenotypes, and disorders. Twin research is considered a key tool in behavioural genetics and in content fields, from biology to psychology. By this definition we already know the track that we are to take in the studies that we are having and the research that we are conducting. Our focus is in making sure that we can draw conclusions of the genes that are influencing a person’s income, and we are to identify if there is a possibility that we can identify the genes.

With his definition in mind, we are using this twin study so that we are able to establish our theory. We are to do an analysis of the data that we are presented with, and the data is the income data. Using RStudio for the analysis, these are the steps that we are to involve In for the analysis. We establish the hereditary (of BMI) maybe in the first one. This can be made possible when we install the package the R and RStudio. Thereafter we are to make sure we have the necessary packages that will make us achieve the analysis. Here we install ‘mets’, and get the data twin2 that is labelled as the ‘income.R’ data. We then begin the analysis of the data so that we are able to make the conclusions that are needed in the research.

The logic of the twin method was simple. [Identical twins](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/monozygotic-twins) have the same genes and share the same parents. Any differences within pairs of MZ twins must be because of their unique environmental experiences. Non-identical twins have the same parents but different genes. Differences within pairs of non-identical twins must be because of the [genetic effects](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/genetic-determinism) and their unique environmental experiences. The conventional approach to the analysis of twin data addresses two issues: (1) do genetic factors contribute significantly to individual differences and (2) what are the relative contributions of genetic and environmental effects to variation within twin pairs? These two questions deal with the related statistical issues of hypothesis testing and estimation. Traditional analyses of twin data focus only on estimating and testing the significance of the genetic component.

After this overview of the analysis the next step is to come up with the models that we are to use so that we are able to make conclusions of the analysis. We have here tried to make an analysis that will help us to ensure that the data we have of the twins is well analyzed and that we are to come up with the graphs that are needed for the visualization of the data. This is made available in the RStudio. We also make sure that the analysis is well placed and that the models that are made here are well stipulated to make sure that there is a well-organized analysis. In the models we are to make use of the regression analysis and classification just to make sure that there is well-placed data in the analysis.

Below is the summary of the statistical data:



GREML

Scores from different cognitive types and abilities are positively correlated and the variance that is shared between the tests is usually termed general intelligence. In this kind of study, we are giving the intelligence to relate positively to the state of the economy of someone. This is because we are generally assuming that the two are positively correlated. All the traits that are portably heritable, but have also been linked.

Firstly, using a recently developed analytic design for combined pedigree and genome-wide molecular genetic data, which in this case we are considering the data we have in the twin2 dataset., we test whether rare genetic variants, CNVs, and structural variants make an additional contribution to the genetic variance in intelligence, neuroticism, and extraversion. Secondly, using unrelated individuals, and genotype data imputed using the Haplotype Reference Consortium (HRC) data, we use minor allele frequency (MAF) stratified GREML (GREML-MS) to quantify the effect of SNPs with a MAF of ≥ 0.001 to determine if this additional variance can also be recovered based on SNPs alone using imputation.

In looking at the GREML we consider intelligence as well the intelligence. General intelligence has been found to be heritable, this is found in the study of the genes of the offsprings. And when we look at the study that we are basing on, that is the economic status, we are relating this to intelligence as well. with twin and family studies estimating that 50 to 80% of phenotypic variance is due to additive genetic factors, a proportion that increases with age from childhood to adulthood.

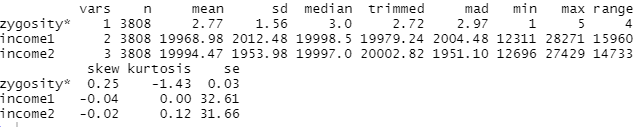
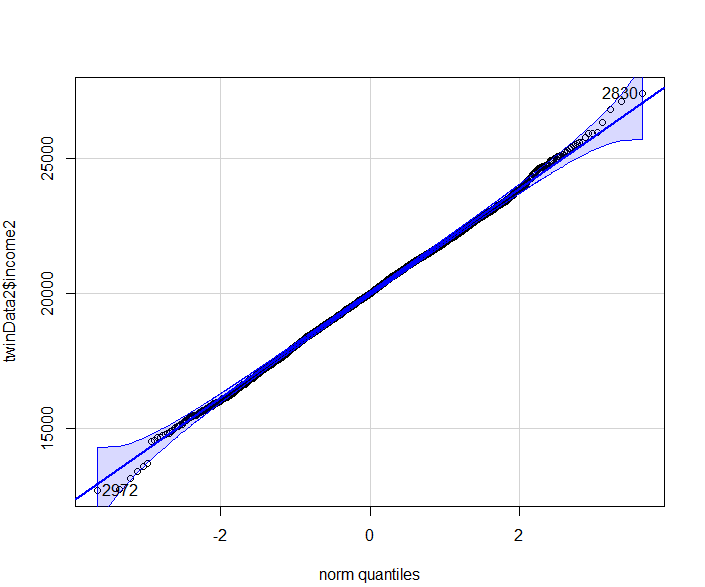
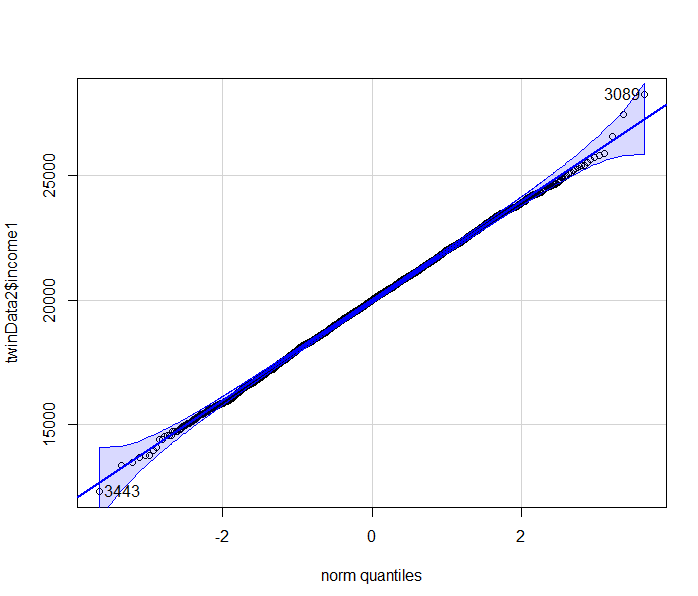
Heritability can also be estimated from molecular genetic data. Using the genomic-relatedness-matrix restricted maximum likelihood single component (GREML-SC) method, the additive effects of common SNPs are estimated to collectively explain between 20 and 50% of variation in general intelligence with an estimate of around 30% in the largest studies. General intelligence is also a significant predictor of fitness components including mortality, fertility and higher social status.

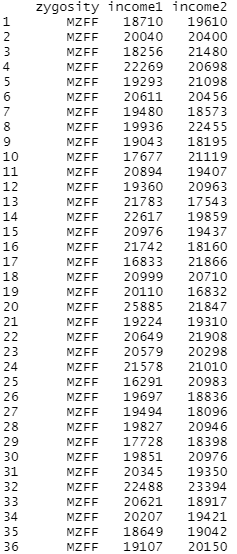
As directional selective pressure on a trait is expected to deplete its genetic variation, the existence of such robust heritability findings seems paradoxical when evolutionary theory is considered . However, mutation-selection balance provides an explanation of how genetic variation can be maintained for quantitative traits that are under directional selective pressure. Mutation-selection balance describes instances where mutations that are deleterious to the phenotype occur within a population at the same rate that they are removed through the effects of selective pressure. Due to the removal of variants with deleterious effects on the phenotype, the existence of common variants with medium to large effects is not expected under mutation-selection balance. This is consistent with the current findings from large genome-wide association studies (GWAS) on cognitive phenotypes, including general intelligence and education, where common SNPs collectively explain a substantial proportion of phenotypic variance, but the individual effect size of each genome-wide significant SNP discovered so far is around 0.02%.

Population genetic simulations show that very rare (MAF < 0.1%) variants explain little of the population variance in traits that are not under selection. However, the contribution made by rare variants increases when their effects on a trait and on fitness are correlated either through pleiotropy, or by the trait directly affecting fitness. The genetically informative evidence that is available tends to show that variants associated with intelligence are also linked to better health, although these effects may be outweighed by a negative effect on fertility. There is also evidence that the regions of the genome making the greatest contribution to intelligence differences have undergone purifying selection. Whereas this does not necessarily imply that intelligence has been selected for or against across our evolutionary history, it does indicate that genetic variants that are associated with intelligence are also associated with fitness, which suggests that rare genetic variants and hence mutation-selection balance, may act to maintain intelligence differences. As with intelligence, heritability estimates for extraversion and neuroticism are much higher, around 34–48%, when based on quantitative (twin- and family-based) genetic methods compared to molecular genetics estimates (4–15% for neuroticism and 0–18% for extraversion. Both extraversion and neuroticism are predictive of social and behavioural outcomes as well as anxiety, well-being, and fertility. Positive genetic correlations have been reported for extraversion with attention deficit hyperactivity disorder and bipolar disorder, and for neuroticism with depression and anorexia nervosa.

In the current study, we quantify the total genetic effect across the autosomes on intelligence (including education, which shows strong genetic correlations with general intelligence and is used as a proxy-phenotype for it in genetic studies), extraversion and neuroticism. Two recent approaches allow us to include genetic variation not normally captured using GWAS. Firstly, as our sample included nominally unrelated individuals with varying degrees of genetic similarity, as well as family members who all provided genome-wide SNP data, we were able to decompose two genetic sources of variance corresponding to genetic effects associated with common SNPs at the population level (h2g), and genetic effects associated with kinship (h2kin) (i.e., associated with SNPs on a family basis). Among related individuals, LD is stronger and hence allows us to capture variation not tagged by common SNPs at the population level. This includes rare variants, CNVs, and other structural variants. As the inclusion of family members can introduce confounding between shared genetic effects and shared environmental effects, we use the GREML-KIN method by Xia and colleagues to control for sibling effects, spouse effects and family effects. By using information from both nuclear family relationships and the many more distant pedigree relationships in the cohort we analyse, this novel approach allows us to estimate kin-specific genetic variation net of common environmental effects. Secondly, we validate the findings for intelligence and education using unrelated individuals by using genotypes imputed using the HRC panel. By using GREML-MS to derive a heritability estimate we were able to include rare SNPs (MAF 0.001–0.01) as well as partition the SNPs by MAF to determine the contribution made to trait variation by rare variants.

Here are the graphs of the income and the qqplots:





Section 4: Feasibility.

The aim of genome‐wide association studies (GWAS) is to identify single nucleotide polymorphisms of which the allele frequencies vary systematically as a function of phenotypic trait values (e.g., between cases with schizophrenia and healthy controls, or between individuals with high vs. low scores on neuroticism). Identification of trait‐associated SNPs may subsequently reveal new insights into the biological mechanisms underlying these phenotypes. Technological advancements allow investigation of the impact of large numbers of SNPs distributed throughout the genome.

To get a clear understanding of GWAS, we need first of all to understand these:

· Clumping: This is a procedure in which only the most significant SNP (i.e., lowest p value) in each LD block is identified and selected for further analyses. This reduces the correlation between the remaining SNPs, while retaining SNPs with the strongest statistical evidence.

· Co‐heritability: This is a measure of the genetic relationship between disorders. The SNP‐based co‐heritability is the proportion of covariance between disorder pairs (e.g., schizophrenia and bipolar disorder) that is explained by SNPs.

· Gene: This is a sequence of nucleotides in the DNA that codes for a molecule (e.g., a protein)

· Heterozygosity: This is the carrying of two different alleles of a specific SNP. The heterozygosity rate of an individual is the proportion of heterozygous genotypes. High levels of heterozygosity within an individual might be an indication of low sample quality whereas low levels of heterozygosity may be due to inbreeding.

· Individual‐level missingness: This is the number of SNPs that are missing for a specific individual. High levels of missingness can be an indication of poor DNA quality or technical problems.

· Linkage disequilibrium (LD): This is a measure of non‐random association between alleles at different loci at the same chromosome in a given population. SNPs are in LD when the frequency of association of their alleles is higher than expected under random assortment. LD concerns patterns of correlations between SNPs.

· Minor allele frequency (MAF): This is the frequency of the least often occurring allele at a specific location. Most studies are underpowered to detect associations with SNPs with a low MAF and therefore exclude these SNPs.

· Population stratification: This is the presence of multiple subpopulations (e.g., individuals with different ethnic background) in a study. Because allele frequencies can differ between subpopulations, population stratification can lead to false positive associations and/or mask true associations. An excellent example of this is the chopstick gene, where a SNP, due to population stratification, accounted for nearly half of the variance in the capacity to eat with chopsticks (Hamer & Sirota, [2000](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0022)).

· Pruning: This is a method to select a subset of markers that are in approximate linkage equilibrium. In PLINK, this method uses the strength of LD between SNPs within a specific window (region) of the chromosome and selects only SNPs that are approximately uncorrelated, based on a user‐specified threshold of LD. In contrast to clumping, pruning does not take the p value of a SNP into account.

· Relatedness: This indicates how strongly a pair of individuals is genetically related. A conventional GWAS assumes that all subjects are unrelated (i.e., no pair of individuals is more closely related than second‐degree relatives). Without appropriate correction, the inclusion of relatives could lead to biased estimations of standard errors of SNP effect sizes. Note that specific tools for analysing family data have been developed.

· Sex discrepancy: This is the difference between the assigned sex and the sex determined based on the genotype. A discrepancy likely points to sample mix‐ups in the lab. Note, this test can only be conducted when SNPs on the sex chromosomes (X and Y) have been assessed.

· Single nucleotide polymorphism (SNP): This is a variation in a single nucleotide (i.e., A, C, G, or T) that occurs at a specific position in the genome. A SNP usually exists as two different forms (e.g., A vs. T). These different forms are called alleles. A SNP with two alleles has three different genotypes (e.g., AA, AT, and TT).

· SNP‐heritability: This is the fraction of phenotypic variance of a trait explained by all SNPs in the analysis.

· SNP‐level missingness: This is the number of individuals in the sample for whom information on a specific SNP is missing. SNPs with a high level of missingness can potentially lead to bias.

· Summary statistics: These are the results obtained after conducting a GWAS, including information on chromosome number, position of the SNP, SNP(rs)‐identifier, MAF, effect size (odds ratio/beta), standard error, and p value. Summary statistics of GWAS are often freely accessible or shared between researchers.

· The Hardy–Weinberg (dis)equilibrium (HWE) law: This concerns the relation between the allele and genotype frequencies. It assumes an indefinitely large population, with no selection, mutation, or migration. The law states that the genotype and the allele frequencies are constant over generations. Violation of the HWE law indicates that genotype frequencies are significantly different from expectations (e.g., if the frequency of allele A = 0.20 and the frequency of allele T = 0.80; the expected frequency of genotype AT is 2\*0.2\*0.8 = 0.32) and the observed frequency should not be significantly different. In GWAS, it is generally assumed that deviations from HWE are the result of genotyping errors. The HWE thresholds in cases are often less stringent than those in controls, as the violation of the HWE law in cases can be indicative of true genetic association with disease risk.

Section 5: Desirability of a GWAS

Genome-wide association studies (GWAS) have been unable to identify variants linked to depression. We hypothesized that examining depressive symptoms and considering gene-environment interaction (G×E) might improve efficiency for gene discovery. We therefore conducted a GWAS and genome-wide environment interaction study (GWEIS) of depressive symptoms.

A basic understanding of the theory behind genetic analysis using GWAS, the essential QC steps, and the use of appropriate software and methods, along with practical experience are imperative to be able to conduct a genetic study with reliable and reproducible results. In this research, it is highlighted the important concepts to successfully conduct a GWAS analysis.

As a GWAS is usually undertaken to increase our understanding of the biological mechanisms that contribute to disease risk, a GWAS will usually be followed up by post‐GWAS analyses. Valuable insights can be acquired by using tools and resources, which enable the researcher to interpret the association results from a functional or biological perspective. GTEx provides information on the association between SNPs and gene expression (Ardlie et al., [2015](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0004)). Ensembl (Birney et al., [2004](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0007)) and FUMA (Watanabe, Taskesen, van Bochoven, & Posthuma, [2017](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0050)) are commonly used tools for functional annotation. In addition, methods that provide important insights into the genetic architecture of the psychiatric trait or disease under study are freely available. For example, GCTA (Yang, Lee, Goddard, & Visscher, [2011](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0053)) and LD score regression analysis (Bulik‐Sullivan et al., [2015](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0008)) have been applied to estimate SNP‐based heritability. Gene‐based tests, which consider the association between a phenotypic trait and multiple SNPs within a gene, (e.g., de Leeuw, Neale, Heskes, & Posthuma, [2016](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0010)) and pathway/gene‐set analyses (de Leeuw et al., [2016](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/#mpr1608-bib-0010)) have increased our insight into the biological pathways of psychiatric disorders. It should be noted that many of the aforementioned methods can be applied using summary statistics. It is beyond the scope of this research to discuss all available post‐GWAS tools and resources in detail.

Here we are able to identify that there are factors that are well linked to the brain, and the diseases, that are eventually linked to one being able to be socially and economically stable. As we analysed the data we looked that in deed the data that we are given cleary indicated that there is a feature that relate to the economic conditions and these are brough out clearly in the state of the mind that one had.

Although in principle all common SNPs could be used in a PRS analysis, it is customary to first clump (see clumping) the GWAS results before computing risk scores. p value thresholds are typically used to remove SNPs that show little or no statistical evidence for association (e.g., only keep SNPs with p values <0.5 or <0.1. Usually, multiple PRS analyses will be performed, with varying thresholds for the p values.

Genome-wide association studies (GWAS) test hundreds of thousands of genetic variants across many genomes to find those statistically associated with a specific trait or disease. This methodology has generated a myriad of robust associations for a range of traits and diseases, and the number of associated variants is expected to grow steadily as GWAS sample sizes increase. GWAS results have a range of applications, such as gaining insight into a phenotype’s underlying biology, estimating its heritability, calculating genetic correlations, making clinical risk predictions, informing drug development programmes and inferring potential causal relationships between risk factors and health outcomes. In this Primer, we provide the reader with an introduction to GWAS, explaining their statistical basis and how they are conducted, describe state-of-the art approaches and discuss limitations and challenges, concluding with an overview of the current and future applications for GWAS results.

Following the publication of the first GWAS 15 years ago, an impressive number of trait-associated variants have been revealed, along with important insights into biology. Current trends in GWAS include an increasingly interdisciplinary approach, covering statistics, data science, genetics and molecular biology. As sample sizes reach more than 1 million participants and genotyping and sequencing costs reduce, GWAS are increasingly using WES and WGS to allow the identification of rare variants, which could potentially explain much of the missing heritability in complex traits, (however, see ref. for a discussion of potential methodological issues in. Minimal phenotyping may be a cost-effective and quick way of gaining power and deep phenotyping and item-level analyses are becoming important to further our understanding of distinct symptoms as opposed to diagnoses, which tend to be a collection of symptoms. Finally, the GWAS field is expanding to better represent the global community through the inclusion of under-represented populations.

Section 6 Conclusion:

Socioeconomic status—essentially, how much money someone earns and what [education](https://www.psychologytoday.com/us/basics/education) and job they have—is an important risk factor for both physical and mental health ([Calixto & Anaya, 2014](https://www.sciencedirect.com/science/article/abs/pii/S1568997214000044?via%3Dihub); [De Vries, 2019](https://www.cambridge.org/core/journals/epidemiology-and-psychiatric-sciences/article/relationship-between-mental-disorders-and-actual-and-desired-subjective-social-status/B790F13864586A3C80D76C2A63457690)). People with lower socioeconomic status have a higher chance of suffering from mental illness during their lifetime than those with higher socioeconomic status ([De Vries, 2019](https://www.cambridge.org/core/journals/epidemiology-and-psychiatric-sciences/article/relationship-between-mental-disorders-and-actual-and-desired-subjective-social-status/B790F13864586A3C80D76C2A63457690)).

So which factors are responsible for the link between socioeconomic status and health?

First and foremost, environmental factors come to mind. These include, for example, better access to healthcare for people with more income, less exposure to potentially harmful environments and better access to healthy food. Interestingly, it has also been suggested that [genes](https://www.psychologytoday.com/us/basics/genetics) can also play a role in the relationship between socioeconomic status and health—for example, if someone has a genetic predisposition for a severe disease, this person might earn less money in the long run in part because they have more sick days. Therefore, even if it sounds somewhat far-fetched, scientists have tried to determine which genetic factors are associated with differences in income.

In a new study published in the scientific journal Nature Communications, the authors investigated genetic factors associated with income in a large sample of 286,301 participants ([Hill et al., 2019](https://www.nature.com/articles/s41467-019-13585-5)). The authors carried out what is called a genome-wide association study (GWAS). GWAS is a technique that is commonly used in genetic research to associate genetic variation to particular variations in the brain and behaviour. Participants give a DNA sample (e.g. a saliva, oral mucosa or blood sample) and DNA is extracted from the sample. Scientists then analyse the genome in these samples by looking at hundreds of thousands or even millions of so-called SNPs (single nucleotide polymorphisms).

These tiny variations in the DNA occur in everybody. By testing lots of individuals from different groups (e.g. different household income groups) and comparing these SNPs systematically, scientists than can identify which specific genes are associated with income.

Using this technique, the scientists could identify 149 different genes that are associated with income. Importantly, these genes showed higher expression in the brain than in other body tissues. In particular, the identified genes were linked to serotonergic and GABA-ergic neurons. The neurotransmitters serotonin and GABA have been linked to higher cognitive functions and emotion in the human brain.

Importantly, the authors found out that many of the genes linked with income have previously been associated with [intelligence](https://www.psychologytoday.com/us/basics/intelligence) in other studies. Therefore, differences in intelligence are likely among the causal factors that lead to differences in income.

Based on these findings, the authors highlight that there are no “genes for income” that directly influence how much we earn. The authors propose that genetic variations affect brain structure and in turn intelligence. Intelligence then influences the education someone has access to (smarter people are more likely to be admitted to prestigious universities). Education then, finally, is associated with income, as people with better education tend to earn more money.

That said, only about 2% of the income differences could be accounted for by genetic factors in this particular data set. Genes may partially and indirectly help explain differences in income, but it has to be kept in mind that the majority of income differences are explained by non-genetic factors.

This is the report generated from the code that we wrote and the graphs that are needed for this analysis.

# **Data Profiling Report**

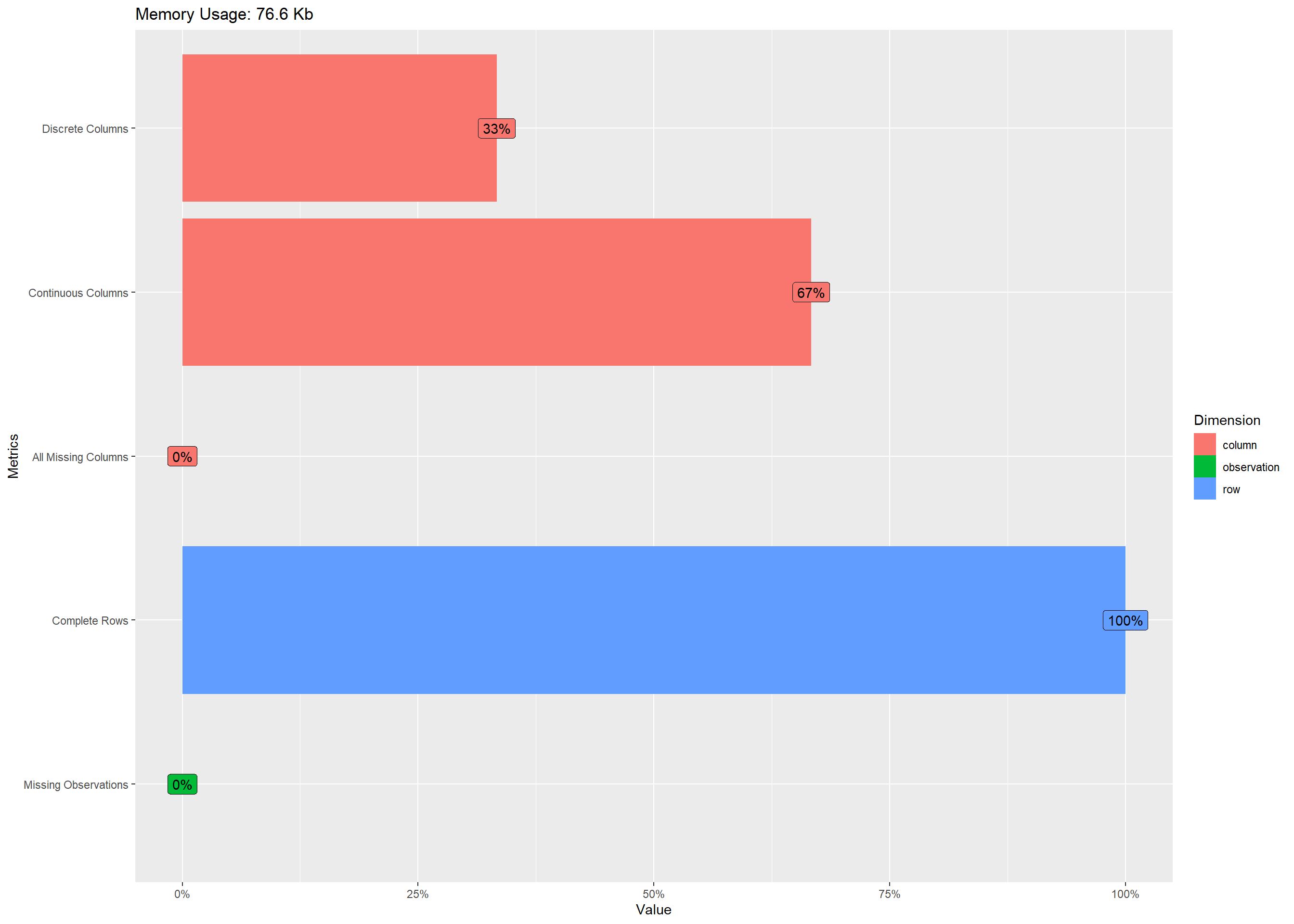
* [Basic Statistics](file:///C:\Users\DESMOND\Documents\report.html#basic-statistics)
  + [Raw Counts](file:///C:\Users\DESMOND\Documents\report.html#raw-counts)
  + [Percentages](file:///C:\Users\DESMOND\Documents\report.html#percentages)
* [Data Structure](file:///C:\Users\DESMOND\Documents\report.html#data-structure)
* [Missing Data Profile](file:///C:\Users\DESMOND\Documents\report.html#missing-data-profile)
* [Univariate Distribution](file:///C:\Users\DESMOND\Documents\report.html#univariate-distribution)
  + [Histogram](file:///C:\Users\DESMOND\Documents\report.html#histogram)
  + [Bar Chart (with frequency)](file:///C:\Users\DESMOND\Documents\report.html#bar-chart-with-frequency)
  + [QQ Plot](file:///C:\Users\DESMOND\Documents\report.html#qq-plot)
* [Correlation Analysis](file:///C:\Users\DESMOND\Documents\report.html#correlation-analysis)
* [Principal Component Analysis](file:///C:\Users\DESMOND\Documents\report.html#principal-component-analysis)

### **Basic Statistics**

#### **Raw Counts**

| **Name** | **Value** |
| --- | --- |
| Rows | 3,808 |
| Columns | 3 |
| Discrete columns | 1 |
| Continuous columns | 2 |
| All missing columns | 0 |
| Missing observations | 0 |
| Complete Rows | 3,808 |
| Total observations | 11,424 |
| Memory allocation | 76.6 Kb |

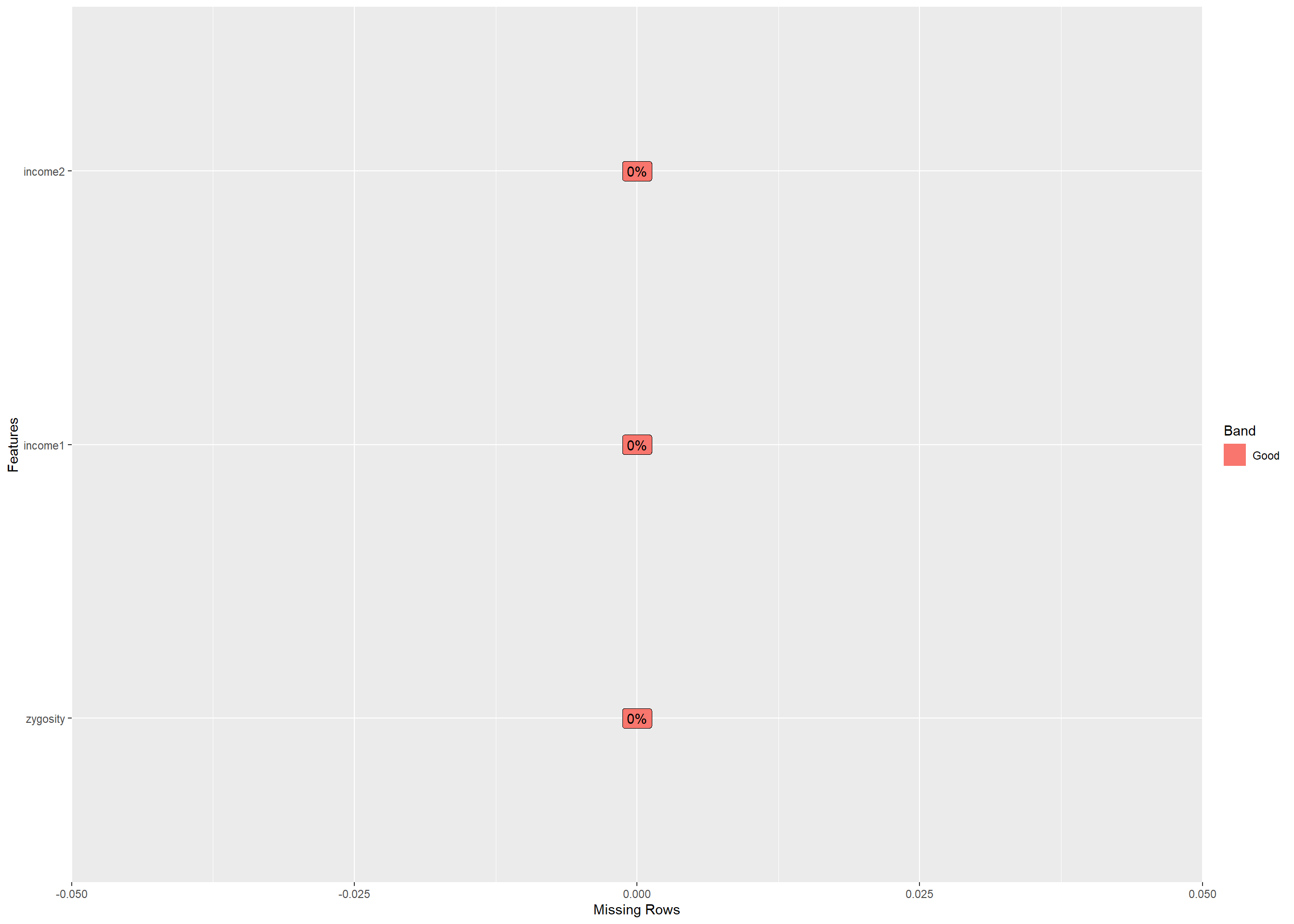
#### **Percentages**



### **Data Structure**

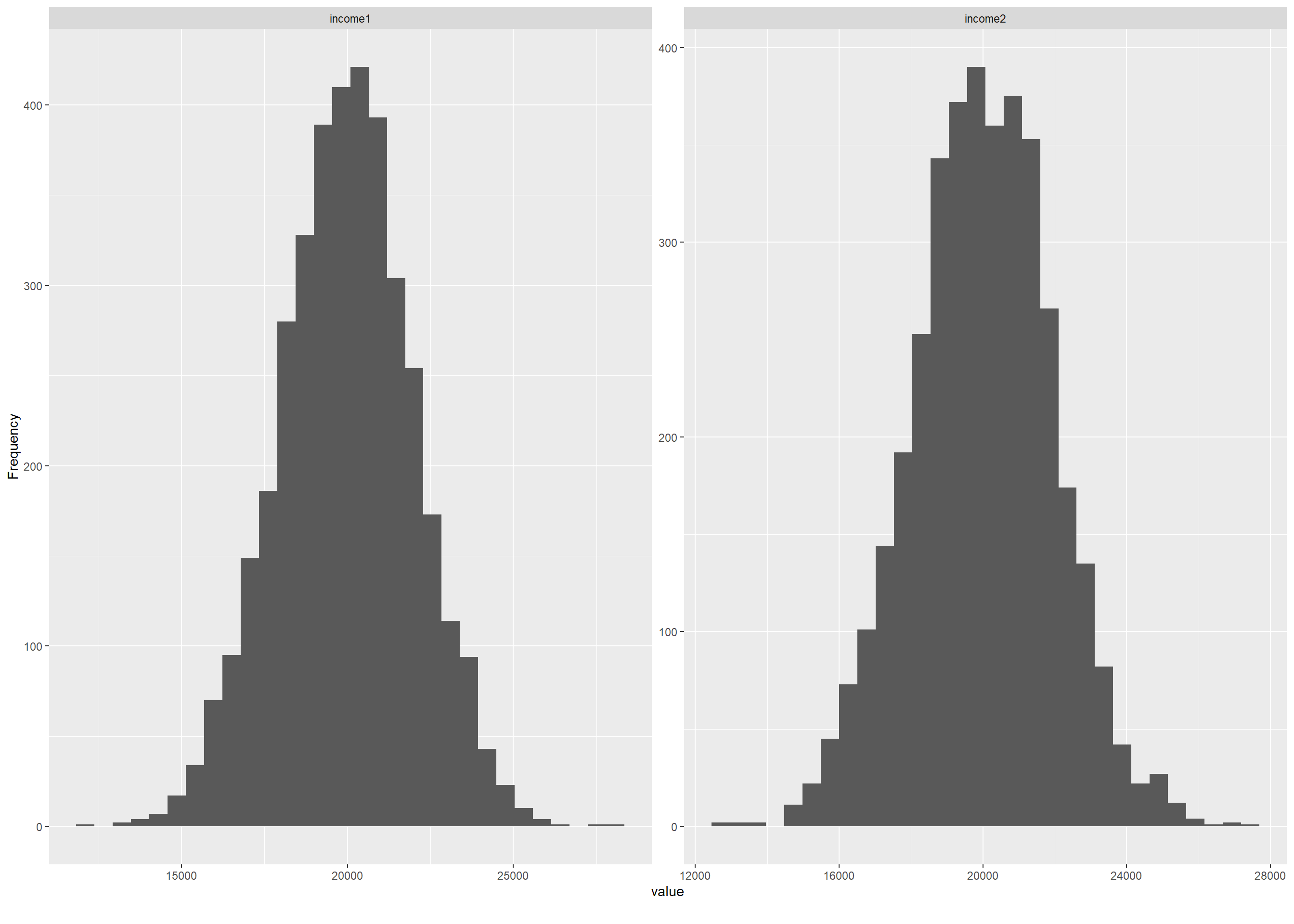
root (Classes 'data.table' and 'data.frame': 3808 obs. of 3 variables:)zygosity (Factor w/ 5 levels "MZFF","MZMM",)income1 (num)income2 (num)

### **Missing Data Profile**

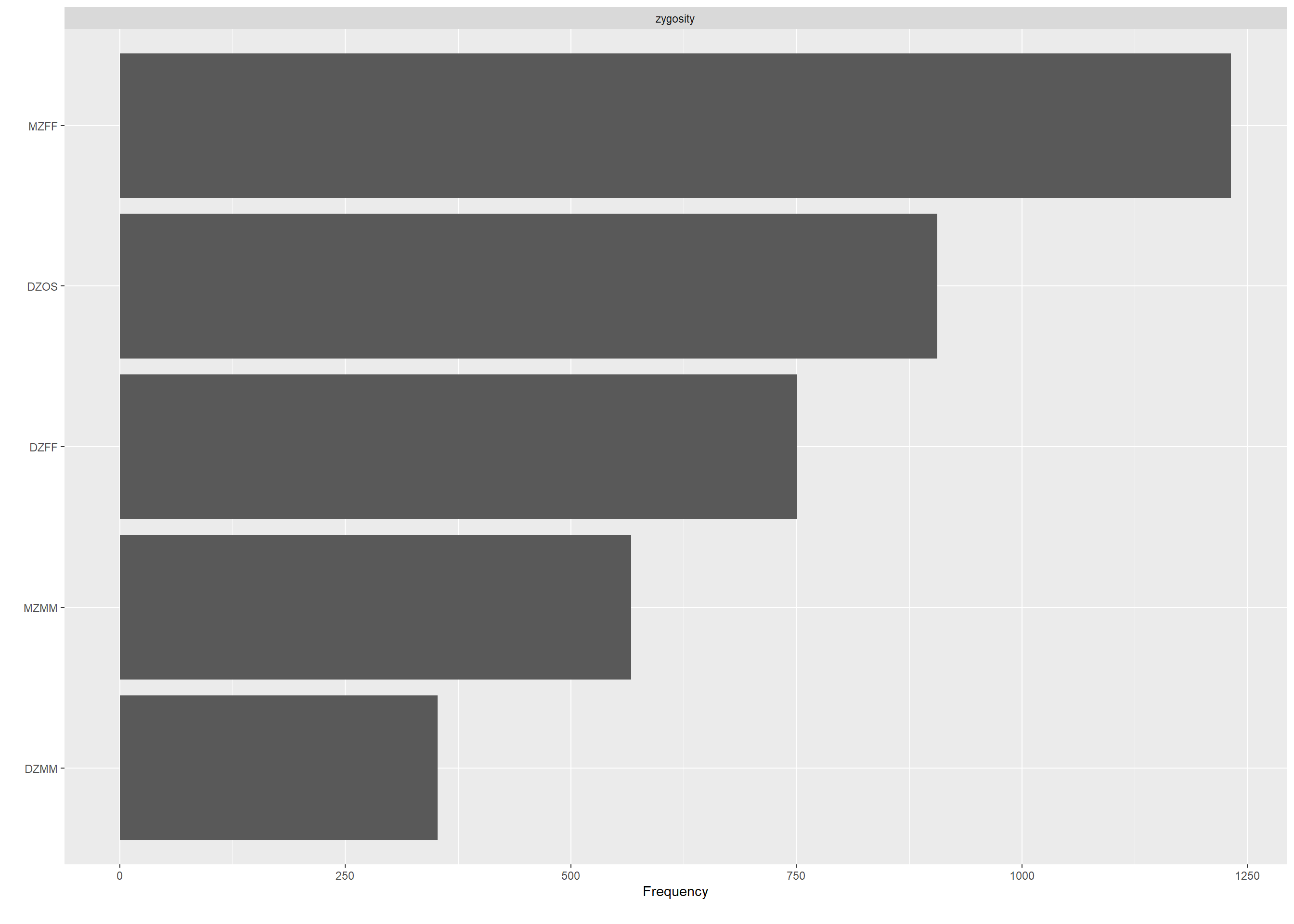


### **Univariate Distribution**

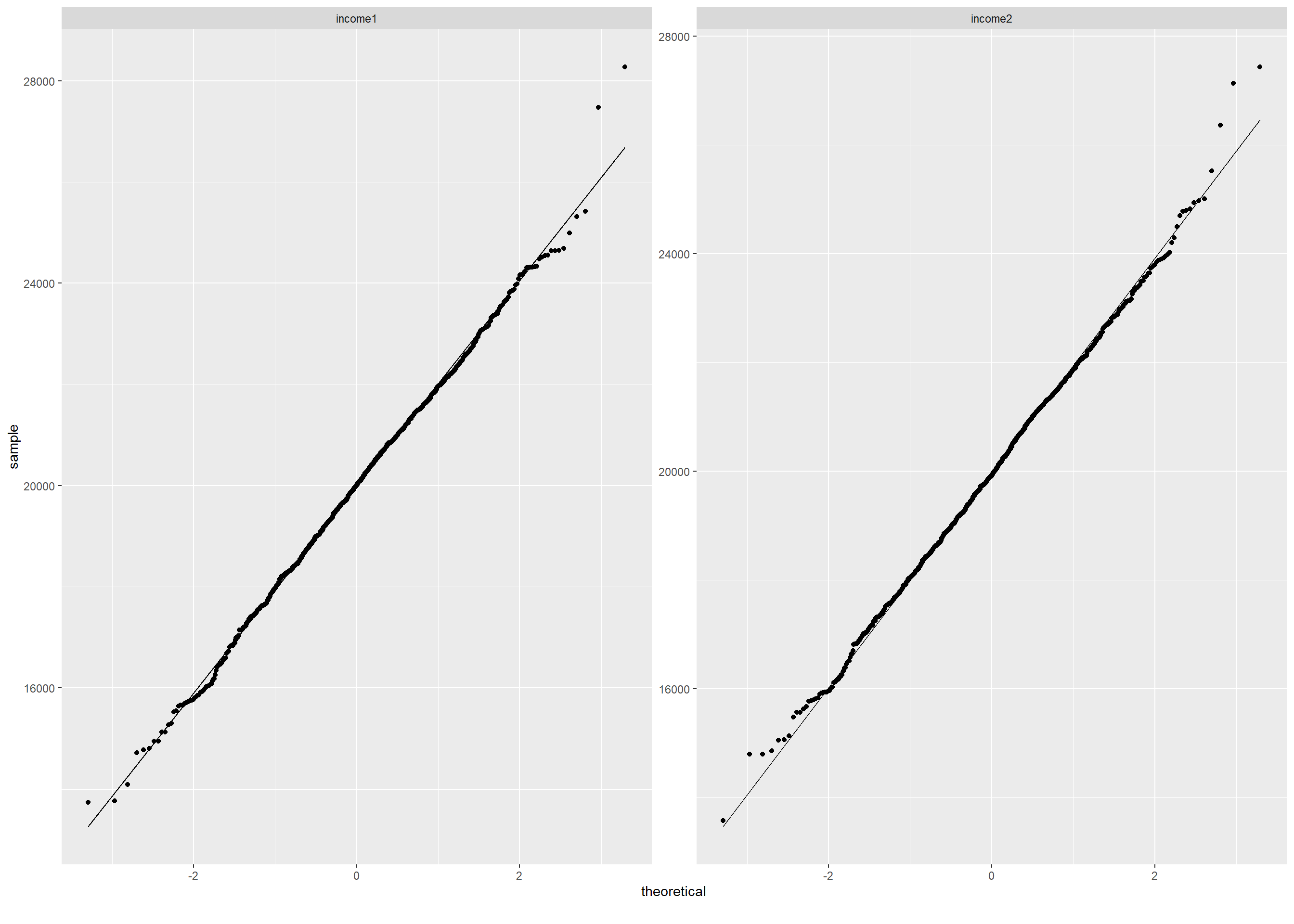
#### **Histogram**



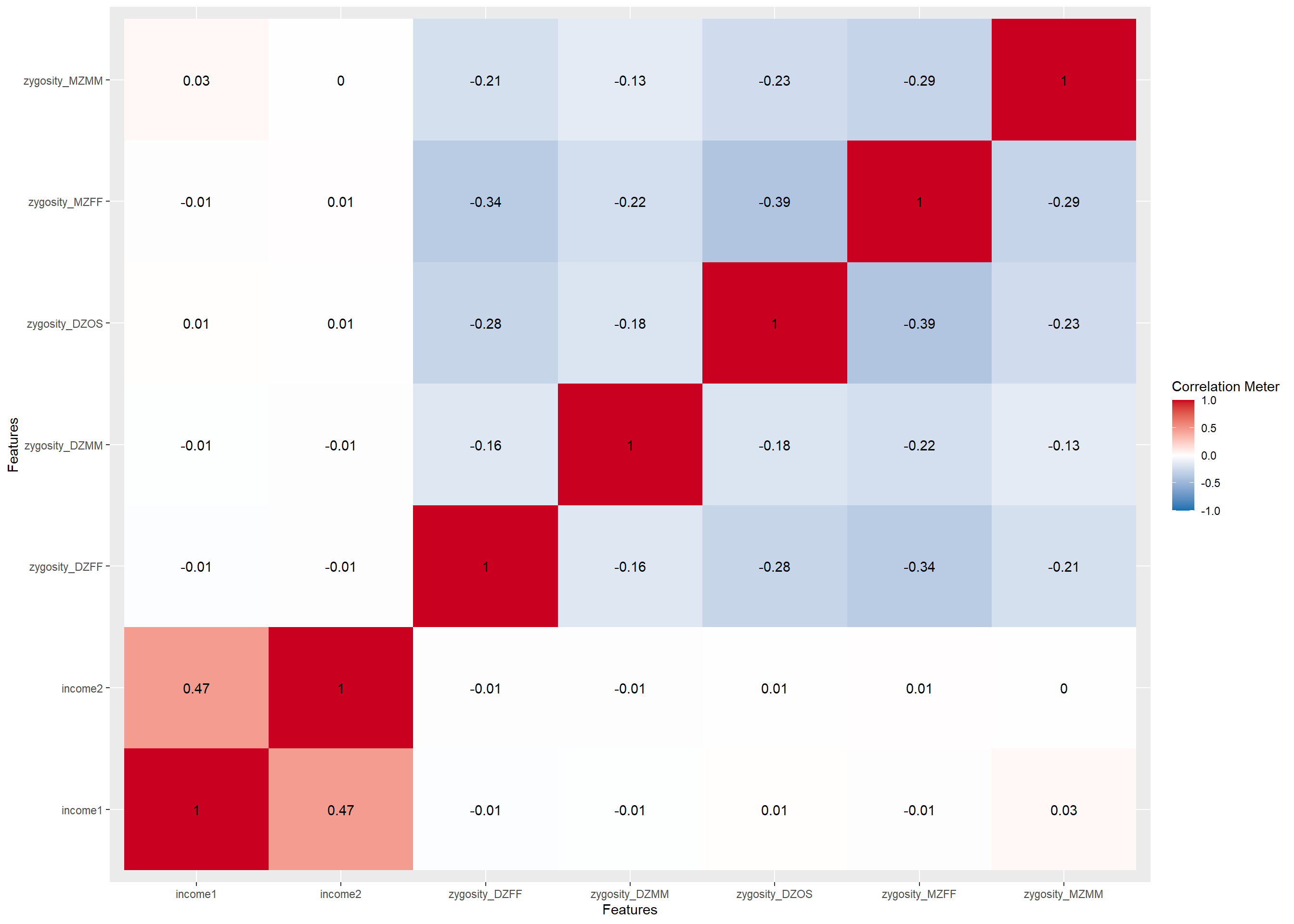
#### **Bar Chart (with frequency)**



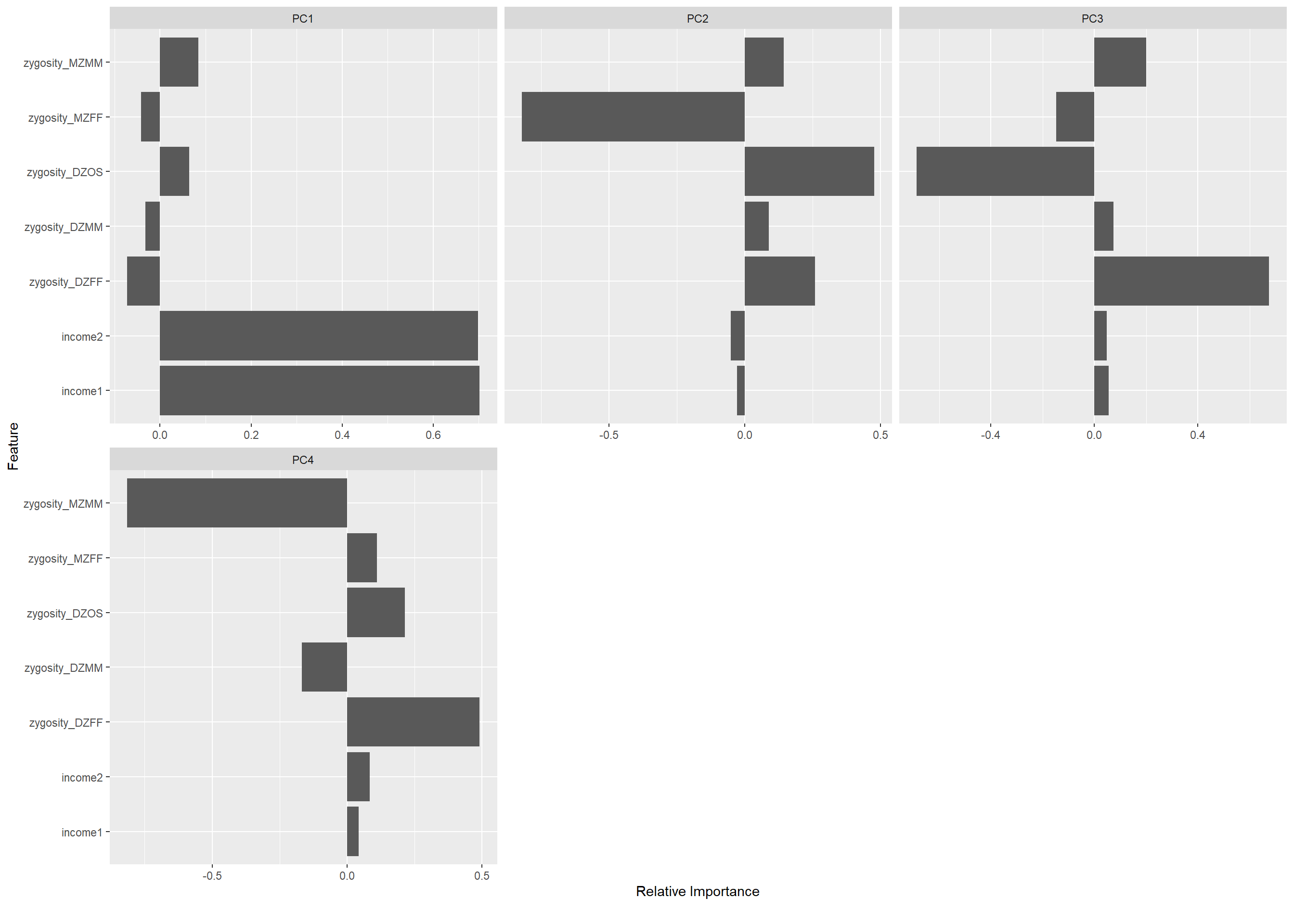
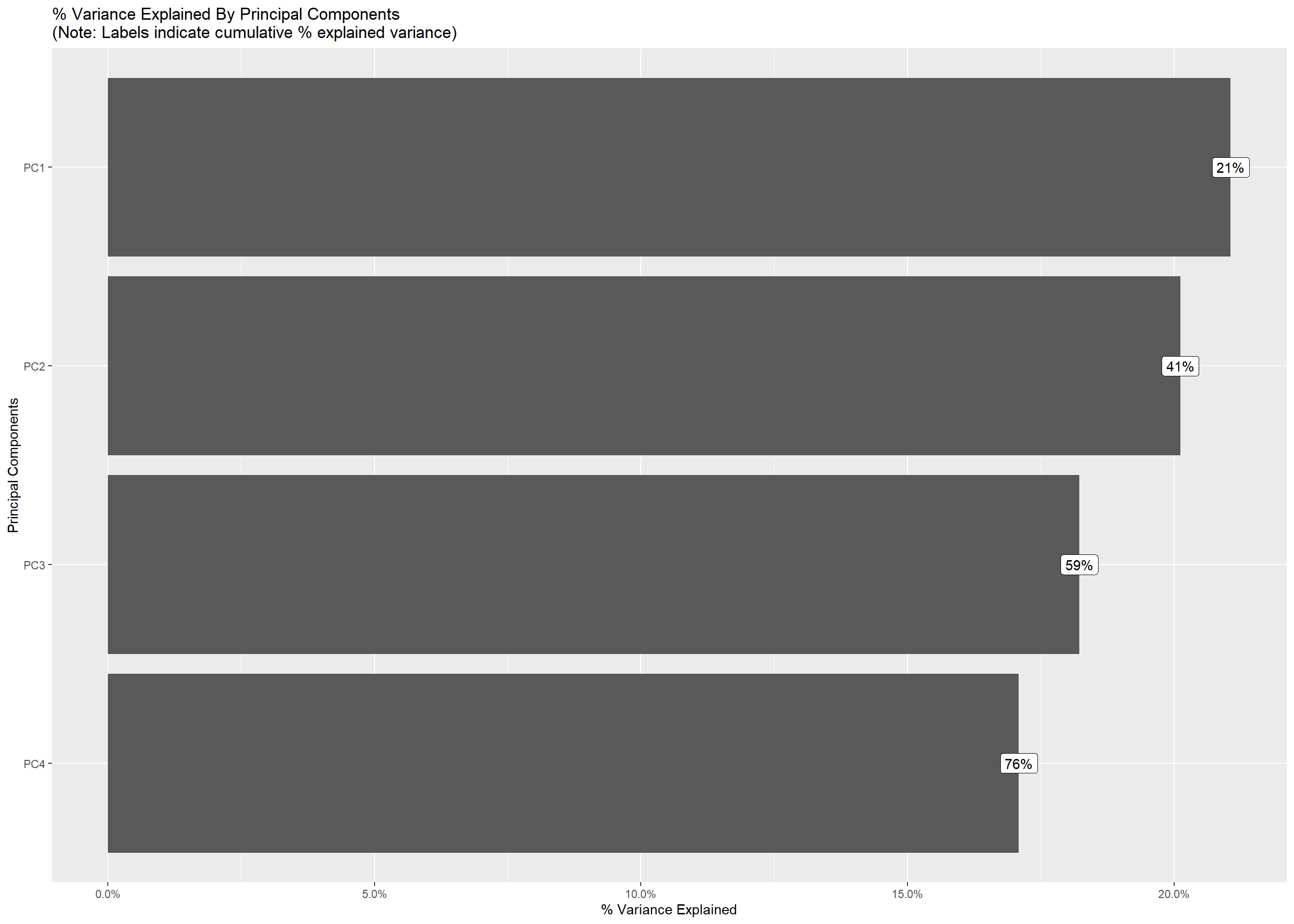
#### **QQ Plot**



### **Correlation Analysis**



### **Principal Component Analysis**



References

*APA PsycNet*. (n.d.). APA PsycNet. <https://psycnet.apa.org/record/2006-05169-011>

*Autistic traits in the general population*. (2003, May 1). JAMA Network | Home of JAMA and the Specialty Journals of the American Medical Association. <https://jamanetwork.com/journals/jamapsychiatry/article-abstract/207441>

(n.d.). CSO - Central Statistics Office. <https://www.cso.ie/en/media/csoie/methods/otherresearch/sdgs/Plotting_choropleth_maps_in_R_with_tmap.pdf>

Data@Urban. (2018, July 10). *How to create state and County maps easily in R*. Medium. <https://urban-institute.medium.com/how-to-create-state-and-county-maps-easily-in-r-577d29300bb2>

*Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood*. (2008, June 2). JAMA Network | Home of JAMA and the Specialty Journals of the American Medical Association. <https://jamanetwork.com/journals/jamapsychiatry/article-abstract/482739>

*Genetics and health - Genes, behavior, and the social environment - NCBI bookshelf*. (n.d.). National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/books/NBK19932/>

*Genome-wide association study (GWAS) and genome-wide environment interaction study (GWEIS) of depressive symptoms in African American and Hispanic/Latina women*. (n.d.). PubMed Central (PMC). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4826276/>

Nast, C. (2019, April 12). *Researchers want to link your genes and income—Should they?* Wired. <https://www.wired.com/story/researchers-want-to-link-your-genes-and-incomeshould-they/>

*Socioeconomic status and the brain: Mechanistic insights from human and animal research*. (2010). Nature Reviews Neuroscience. <https://www.nature.com/articles/nrn2897>

*A tutorial on conducting genome‐wide association studies: Quality control and statistical analysis*. (n.d.). PubMed Central (PMC). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/>

*When to use waterfall vs. Agile*. (2019, May 17). Macadamian. <https://www.macadamian.com/learn/when-to-use-waterfall-vs-agile/>