R Plot

Team 7

Decmeber 13th, 2018

INFM 600

**R PLOT DOCUMENT FOR 600**

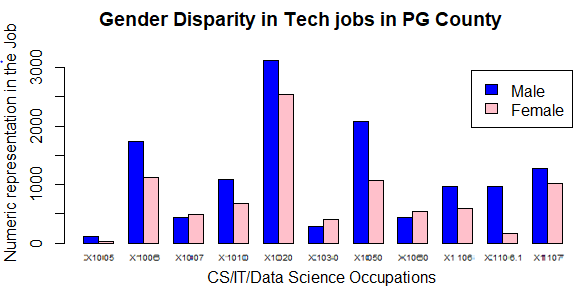
DAN YANG PARTH BORICHA EBANEZARE TADELE

**Distribution of Tech Jobs in Prince George’s County based on Gender**

> library(readr)

> CSjobGender <- read\_csv("C:/Users/Parth Boricha/Downloads/CSjobGender.csv")

> View(CSjobGender)



From the above graph, it is clearly visible that in eight out of eleven tech jobs with focus in CS/IT/Data Science in Prince George’s County more men are employed then women.

Also, we know from our dataset that proportion of men and women for all races in PG County is a nearly 50-50 split.

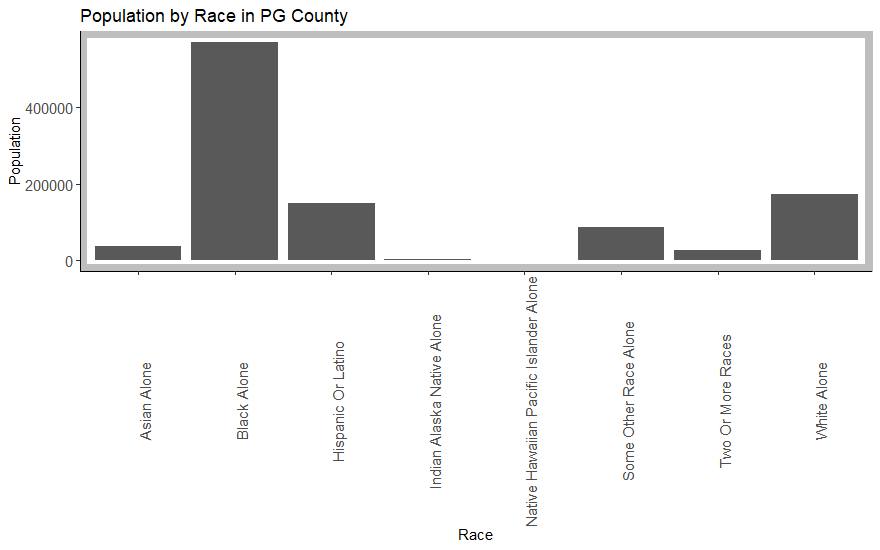
Since our focus have been women of color we can infer from the above graph that women of color are underrepresented in these tech roles and therefore our idea of making accessible bootcamps for them with help of government and NGO’s can be of significant help.

We have made use of ggplot2 to get this visualization. Also, a lot more goes into this because we were asked to include a key/legend. We have therefore added a legend for showing the color code for men and women.

Proper namings of axes and readable text on the graph has been incorporated.

**Population by Race in Prince George’s County**

> ggplot(data = PGCoRace, aes( x = Race, y = Population)) + geom\_bar(stat = "identity", position = position\_dodge()) + xlab("Race") + ylab("Population")+ theme(axis.text.x = element\_text(angle = 90, size = 11), axis.text.y = element\_text(size=11),axis.title.y = element\_text(size=11),axis.title.x = element\_text(size=11),panel.background = element\_blank(),panel.grid.major = element\_blank(),panel.grid.minor = element\_blank(),axis.line = element\_line(colour = "black"),panel.border = element\_rect(colour = "brown", fill=NA, size=5)) + labs(title = "Population by Race in PG County")



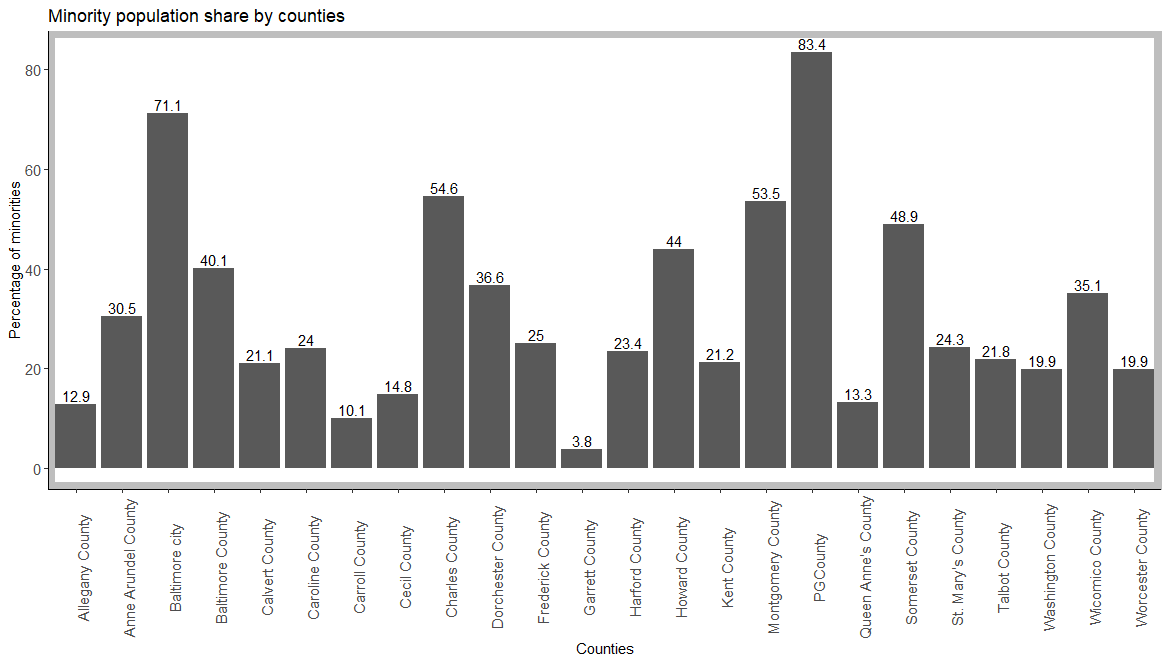
The above bar graph shows us the racial demographic makeup of PG county.

So why did we choose PG County for our analysis?

This graph answers that question. Our area of focus was women of color. An clearly women from minority background make up a majority of the population and therefore helps in our goal of making them have better access to online technical courses and bootcamps.

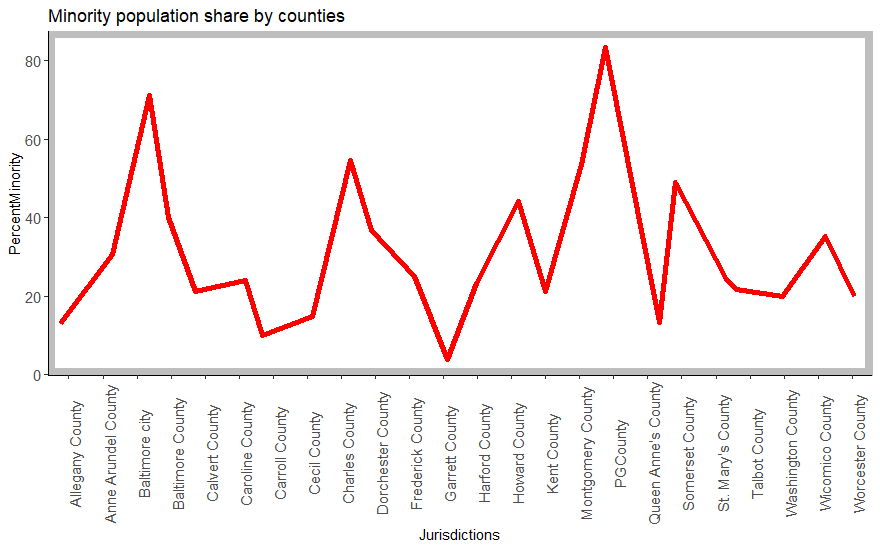
**Minority Share Population by Counties**

ggplot(data = mdSocioEcMinSum, aes( x = Jurisdictions, y = PercentMinority)) + geom\_bar(stat = "identity", position = position\_dodge(width = 2)) + xlab("Counties") + ylab("Percentage of minorities")+ theme(axis.text.x = element\_text(angle = 90, size = 10), axis.text.y = element\_text(size=10),axis.title.y = element\_text(size=10),axis.title.x = element\_text(size=10),panel.background = element\_blank(),panel.grid.major = element\_blank(),panel.grid.minor = element\_blank(),axis.line = element\_line(colour = "black"),panel.border = element\_rect(colour = "gray", fill=NA, size=5)) + labs(title = "Minority population share by counties")+geom\_text(aes(label=PercentMinority), position=position\_dodge(width=0.9), vjust=-0.25)



Here we are comparing the percentage minority population of various counties to highlight were PG County stands in comparison to others. It is clear that PG county is by far the most diverse in terms of minority percentage in overall population at a staggering 83.4%. The next closest comes at 71.1% which is more than 10 points less than PG County. Other counties have even fewer minority populations. Once again, this bar graph reaffirms our decision to choose PG County as our county of focus.

> ggplot(data = mdSocioEcMinSum, aes( x = Jurisdictions, y = PercentMinority, group = 1)) + geom\_line(position = "jitter", colour = "red", size = 2) + theme(axis.text.x = element\_text(angle = 90, size = 11), axis.text.y = element\_text(size=11),axis.title.y = element\_text(size=11),axis.title.x = element\_text(size=11),panel.background = element\_blank(),panel.grid.major = element\_blank(),panel.grid.minor = element\_blank(),axis.line = element\_line(colour = "black"),panel.border = element\_rect(colour = "gray", fill=NA, size=5)) + labs(title = "Minority population share by counties") + xkcd\_theme()



We came with this graph in order to make usage of XKCD package which helps us give that jittery red line graph effect which feels like it is probably drawn in hand instead of having a normal ggplot2 generated line. I also intended to use the humor font using XKCD but couldn’t install the font on my computer due to compatibility issues.

> # here we want to see the relationship between unemployed and family in provity

> x2<- mdSocioEc$Unemployment.Rate

> y2 <- mdSocioEc$PercentFamiliesinPoverty

> relation2 <- lm(y2~x2)

> lm(formula = y2~x2)

Call:

lm(formula = y2 ~ x2)

Coefficients:

(Intercept) x2

-4.395 1.851

> summary(lm(y2~x2))

Call:

lm(formula = y2 ~ x2)

Residuals:

Min 1Q Median 3Q Max

-6.6392 -1.5311 -0.5254 0.8640 9.4505

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -4.3953 2.2947 -1.915 0.0685 .

x2 1.8506 0.3305 5.600 1.25e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

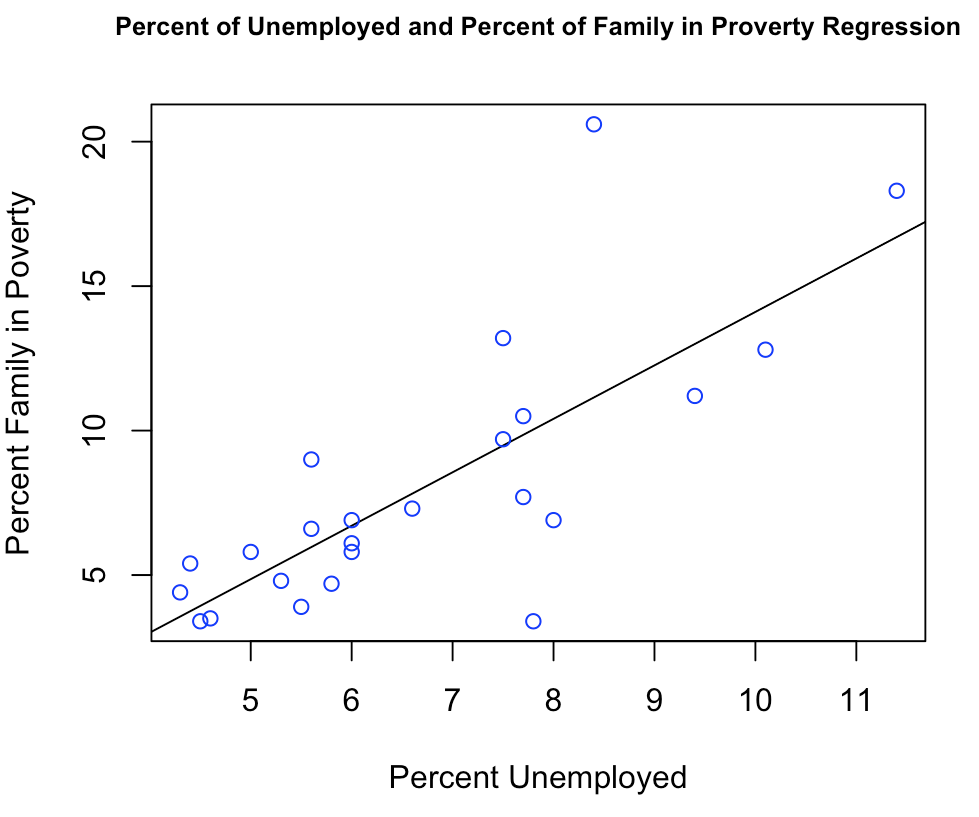
Residual standard error: 2.98 on 22 degrees of freedom

Multiple R-squared: 0.5877, Adjusted R-squared: 0.569

F-statistic: 31.36 on 1 and 22 DF, p-value: 1.249e-05

> plot(x2, y2, col="blue", main = "Percent of Unemployed and Percent of Family in Proverty Regression", abline(lm(y2~x2)),cex.main = 0.8, xlab = "Percent Unemployed", ylab= "Percent Family in Poverty")

Since the P value is very small, so the X2 is very significant.



This graph clearly shows a positive linear correlation between poverty rate and unemployment. Therefore, the idea of helping marginalized group for connecting them with NGO’s who can make Data Science related tech courses available to them for free.

**Online tech couses are already cheaper than non-tech courses?**

We use unique(TA$Major) to get infer that there are 79 different majors people are opting for

aggregate(TA$Cost, list(Major = TA$Major), mean)

 Major         x

1                                             Accounting (Business) 645.2757

2                                             Aerospace Engineering 940.0000

3                                          African-American Studies 366.0000

4                                              Agricultural General 168.0000

5                       Animal Sciences: Animal Care and Management  403.3333

6                                                      Architecture 325.6000

7                                                     Art Education 372.0000

8                                                       Art History 305.0000

9                                                    Bioengineering 1830.0000

10                       Biological Sciences: Ecology and Evolution  354.0000

11                             Biological Sciences: General Biology  686.6667

12                                            Business/Admin./Mgmt. 775.9929

13                   Central European, Russian and Eurasian Studies  482.5000

14                              Civil and Environmental Engineering  452.5500

15                                                    Communication 595.6875

16                                                 Community Health 1343.6667

17                                             Computer Engineering 380.4000

18                                                 Computer Science 674.9217

19                                      Computer Systems Management 588.6538

20                                                 Criminal Justice 697.0933

21                                 Criminology and Criminal Justice 771.2214

22                                        Early Childhood Education 1013.7500

23                                    Education (Teacher:Undecided) 1435.2500

24                                           Electrical Engineering 452.0000

25                                             Emergency Management 654.0522

26                                          Engineering (Undecided) 639.6500

27                                  English Language and Literature 720.9350

28                        Environmental Science and Policy (Policy) 1881.0000

29                       Environmental Science and Policy (Science)  878.0000

30                   Environmental, Educational and Park Management 1077.8571

31                                                   Family Studies 1464.0000

32                                               Finance (Business) 1352.4667

33                                      Fire Protection Engineering 630.6167

34                                                     Fire Science 502.8623

35                                                 Forensic Science 1402.1429

36                                  General Business and Management 1306.6667

37                                                  General Studies 383.8517

38                                                        Geography 535.4250

39                                          Government and Politics 1273.7143

40                                          Health & Human Services 789.3700

41                                                          History 654.3997

42                                                  Human Resources 720.5636

43                                                             HVAC 364.4750

44                                               Individual Studies 735.0000

45                                     Information Systems-Business 1150.6667

46                                           Information Technology 910.1266

47                                           International Business 813.0000

48                                                      Job Related 954.9978

49                                                              Law 1022.2646

50                                        Liberal Arts/Gen. Studies 185.3333

51                                                      Linguistics 577.5000

52 Logistics, Transportation and Supply Chain Management (Business) 1374.0000

53                                             Marketing (Business) 749.0000

54                                                      Mathematics 810.5000

55                                           Mechanical Engineering 1107.4375

56                                                             None 745.8185

57                                                       Not Listed 756.3678

58                                                    Not Specified 686.5805

59                                                          Nursing 746.9164

60                                             Nutritional Sciences 501.0833

61                                 Operations Management (Business) 1041.8000

62                                                      Other/Misc. 769.2516

63                                                Paralegal Studies 442.3429

64                                               Physical Education 386.0000

65                                                Political Science 947.4545

66                                           Pre-Medical Technology 292.0000

67                                                      Pre-Nursing 347.2222

68                                           Professional/Technical 520.7143

69                                                       Psychology 704.3370

70                                            Public Administration 1196.2807

71                                                    Public Safety 976.4076

72                                  Russian Language and Literature 1830.0000

73                                                Science Education 585.0000

74                                         Social Studies Education 1930.0000

75                                                      Social Work 982.5535

76                                                        Sociology 607.9000

77                                  Spanish Language and Literature 909.6860

78                                 Undecided (Letters and Sciences) 240.0000

79                                                   Urban Forestry 528.0000

From this data set we subset the data amongst two excel sheets TA3 & TA4 such that TA3 has all non-computer/data science courses and TA4 has all computer and data science related courses.

Now using R script: mean(TA3$Cost) = 791.5187351

mean(TA4$Cost) = 740.95376

It is astonishing that computer/data science related courses actually cost lesser

This further strengthens our idea of promoting these online computer/data science related courses even if there is no funding from Government or NGOs in case.

WORD COUNT = 1274