Job\_Trend.r

preetha.varma

Mon Feb 22 19:42:49 2016

Salary2011<- read.csv("state-of-california-2011.csv",header = TRUE,strip.white = TRUE)  
Salary2014 <- read.csv("state-of-california-2014.csv",header = TRUE,strip.white = TRUE)  
Salary2012 <- read.csv("state-of-california-2012.csv",header = TRUE,strip.white = TRUE)  
Salary2013 <- read.csv("state-of-california-2013.csv",header = TRUE,strip.white = TRUE)  
  
Salary2011$job\_title<-trimws(Salary2011$job\_title)  
Salary2012$job\_title<-trimws(Salary2012$job\_title)  
Salary2013$Job.Title<-trimws(Salary2013$Job.Title)  
Salary2014$Job.Title<-trimws(Salary2014$Job.Title)  
  
  
  
#Test if all strings are trimmed - The below string had a long white space even after strip.white = true  
#Salary2014 %>% filter(Job.Title =='OFFICER, CALIFORNIA HIGHWAY PATROL')  
#Salary2011 %>% filter(job\_title == 'OFFICER, CALIFORNIA HIGHWAY PATROL')  
  
  
length(Salary2011$job\_title)

## [1] 246599

length(Salary2014$Job.Title)

## [1] 240531

length(Salary2011$job\_title)-length(Salary2014$Job.Title)

## [1] 6068

length(unique(Salary2011$job\_title))

## [1] 4047

length(unique(Salary2012$job\_title))

## [1] 4037

length(unique(Salary2013$Job.Title))

## [1] 3957

length(unique(Salary2014$Job.Title))

## [1] 3923

str(Salary2014$Job.Title)

## chr [1:240531] "CHIEF INVESTMENT OFFICER, PUBLIC EMPLOYEES' RETIREMENT SYSTEM" ...

library(dplyr)

##   
## Attaching package: 'dplyr'

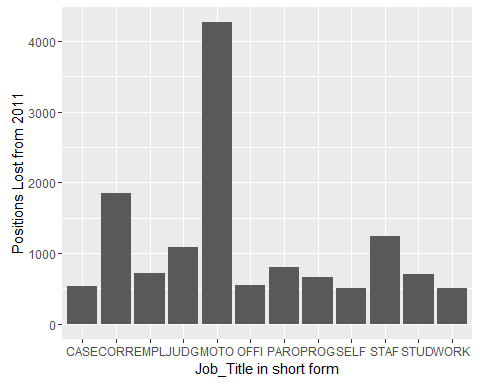
## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

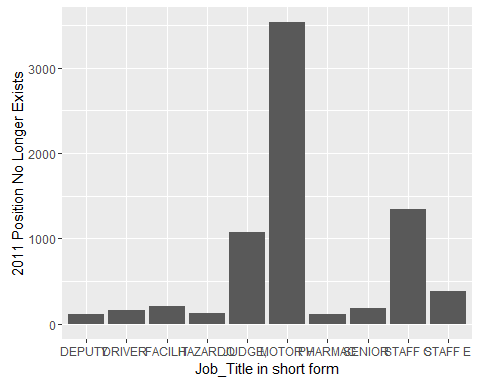
library(ggplot2)  
library(tidyr)  
  
  
#JobTitle2011 <- c(unique(as.character(Salary2011$job\_title)))  
#JobTitle2012 <- c(unique(as.character(Salary2012$job\_title)))  
#JobTitle2013 <- c(unique(as.character(Salary2013$Job.Title)))  
#JobTitle2014 <- c(unique(as.character(Salary2014$Job.Title)))  
  
#summary(Salary2011$job\_title)  
  
# create a dataframe to store 2011 Jobtitle and number of position.  
Jobtitle2011byCount <- data.frame(table(Salary2011$job\_title))  
  
  
names(Jobtitle2011byCount)[1] <- paste("Job\_Title")  
names(Jobtitle2011byCount)[2] <- paste("Number\_Of\_Positions")  
  
# create a dataframe to store 2012 Jobtitle and number of position.  
Jobtitle2012byCount <- data.frame(table(Salary2012$job\_title))  
  
  
names(Jobtitle2012byCount)[1] <- paste("Job\_Title")  
names(Jobtitle2012byCount)[2] <- paste("Number\_Of\_Positions")  
  
# create a dataframe to store 2013 Jobtitle and number of position.  
Jobtitle2013byCount <- data.frame(table(Salary2013$Job.Title))  
  
  
names(Jobtitle2013byCount)[1] <- paste("Job\_Title")  
names(Jobtitle2013byCount)[2] <- paste("Number\_Of\_Positions")  
  
# create a dataframe to store 2014 Jobtitle and number of position.  
Jobtitle2014byCount <- data.frame(table(Salary2014$Job.Title))  
  
  
names(Jobtitle2014byCount)[1] <- paste("Job\_Title")  
names(Jobtitle2014byCount)[2] <- paste("Number\_Of\_Positions")  
  
#Jobtitle2012byCount %>% filter(Job\_Title %in% ('ACCOUNT CLERK II'))  
  
  
#leftjoin2011 <- left\_join(Jobtitle2011byCount,Jobtitle2014byCount,"Job\_Title")  
#names(leftjoin2011)[2] <- paste("2011Job\_Title")  
#names(leftjoin2011)[3] <- paste("2014Job\_Title")  
  
#Identify Job Titles and Position reduced/removed from 2011  
  
#Join all jobtitles of 2011 and 2014  
join20112014 <- full\_join(Jobtitle2011byCount,Jobtitle2014byCount,"Job\_Title")

## Warning in outer\_join\_impl(x, y, by$x, by$y): joining factors with  
## different levels, coercing to character vector

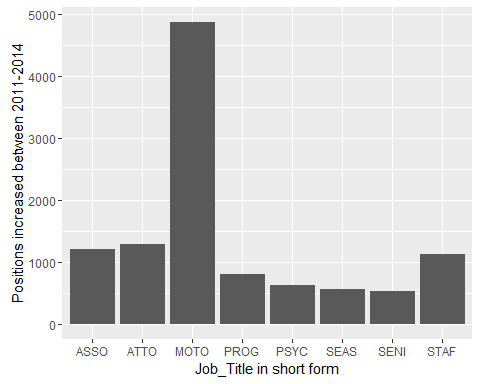
names(join20112014)[2] <- paste("2011Positions")  
names(join20112014)[3] <- paste("2014Positions")  
  
join20112014$`2014Positions`[is.na(join20112014$`2014Positions`)] <- 0  
join20112014$`2011Positions`[is.na(join20112014$`2011Positions`)] <- 0  
  
#Jobs reduced from 2011  
jobsreducedfrom2011 <- join20112014 %>% filter(`2011Positions` - `2014Positions`>10)  
  
MajorJobCutFrom2011 <- join20112014 %>% filter(`2011Positions` - `2014Positions`>500)  
  
MajorJobCutFrom2011$SJT <- strtrim(MajorJobCutFrom2011$Job\_Title, 4)  
MajorJobCutFrom2011$PositionsLost <- MajorJobCutFrom2011$`2011Positions` - MajorJobCutFrom2011$`2014Positions`  
  
#Display count of jobs titles reduced from 2011.  
ggplot(data=MajorJobCutFrom2011, aes(x=SJT, y=PositionsLost)) +  
 geom\_bar(stat="identity")+  
 ylab('Positions Lost from 2011')+  
 xlab('Job\_Title in short form')



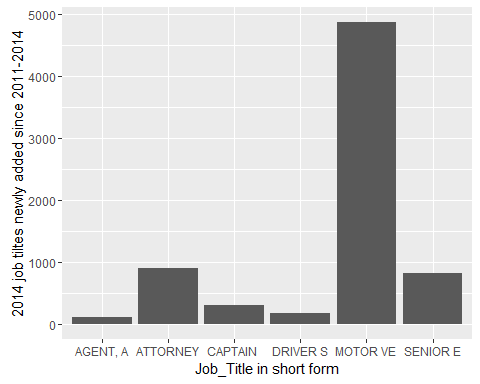
#Jobs titles that no longer exists from 2011  
  
JobTitleNoLongerExists <- join20112014 %>% filter(`2014Positions` == 0)  
  
  
#Jobs titles that no longer exists from 2011 which once had positions more than 100 Positions  
  
MajorJobTitlescut2011 <- join20112014 %>% filter(`2014Positions` == 0 & `2011Positions` >100)  
MajorJobTitlescut2011$SJT <- strtrim(MajorJobTitlescut2011$Job\_Title, 7)  
  
#Display jobs titles removed from 2011.  
ggplot(data=MajorJobTitlescut2011, aes(x=SJT, y=`2011Positions`)) +  
 geom\_bar(stat="identity")+  
 ylab('2011 Position No Longer Exists')+  
 xlab('Job\_Title in short form')



#Identify Job Titles and Position increased/newly added in 2014  
  
#Jobs increased from 2011  
jobsIncreasedfrom2011 <- join20112014 %>% filter(`2014Positions` - `2011Positions`>10)  
  
MajorjobsIncreasedfrom2011 <- join20112014 %>% filter(`2014Positions` - `2011Positions`>500)  
  
  
MajorjobsIncreasedfrom2011$SJT <- strtrim(MajorjobsIncreasedfrom2011$Job\_Title, 4)  
MajorjobsIncreasedfrom2011$PositionsIncreased <- MajorjobsIncreasedfrom2011$`2014Positions` - MajorjobsIncreasedfrom2011$`2011Positions`  
  
#Display jobs titles increased from 2011.  
ggplot(data=MajorjobsIncreasedfrom2011, aes(x=SJT, y=PositionsIncreased)) +  
 geom\_bar(stat="identity")+  
 ylab('Positions increased between 2011-2014')+  
 xlab('Job\_Title in short form')



#Jobs titles newly added after 2011  
  
MajorJobTitlesIncreaedFrom2011 <- join20112014 %>% filter(`2011Positions` == 0 & `2014Positions` >100)  
MajorJobTitlesIncreaedFrom2011$SJT <- strtrim(MajorJobTitlesIncreaedFrom2011$Job\_Title, 8)  
  
#Display new jobs titles added after 2011.  
ggplot(data=MajorJobTitlesIncreaedFrom2011, aes(x=SJT, y=`2014Positions`)) +  
 geom\_bar(stat="identity")+  
 ylab('2014 job tiltes newly added since 2011-2014')+  
 xlab('Job\_Title in short form')



#Job Cut Density  
  
  
  
MajorJobCutFrom2011 <- join20112014 %>% filter(`2011Positions` - `2014Positions`>200)  
MajorJobCutFrom2011$SJT <- strtrim(MajorJobCutFrom2011$Job\_Title, 4)  
MajorJobCutFrom2011$PositionsLost <- MajorJobCutFrom2011$`2011Positions` - MajorJobCutFrom2011$`2014Positions`  
  
  
names(Salary2011)[2]<- paste('Job\_Title')  
names(Salary2014)[2]<- paste('Job\_Title')  
  
JobCutDensity2011 <- left\_join(MajorJobCutFrom2011,Salary2011,"Job\_Title")  
JobCutDensity2011$Status <- 'NA'  
#Ensure Year variable has value 2011  
JobCutDensity2011$year <- 2011  
  
JobCutDensity2014 <- left\_join(MajorJobCutFrom2011,Salary2014,"Job\_Title")  
#Ensure Year variable has value 2014  
JobCutDensity2014$Year <- 2014  
  
glimpse(JobCutDensity2014)

## Observations: 52,970  
## Variables: 16  
## $ Job\_Title (chr) "ACCOUNTING TECHNICIAN", "ACCOUNTING TECH...  
## $ 2011Positions (dbl) 529, 529, 529, 529, 529, 529, 529, 529, 5...  
## $ 2014Positions (dbl) 280, 280, 280, 280, 280, 280, 280, 280, 2...  
## $ SJT (chr) "ACCO", "ACCO", "ACCO", "ACCO", "ACCO", "...  
## $ PositionsLost (dbl) 249, 249, 249, 249, 249, 249, 249, 249, 2...  
## $ Employee.Name (fctr) Pamela J Turknett, Amy C Bell, Ceasar B ...  
## $ Base.Pay (dbl) 34756.32, 41372.44, 56923.68, 57503.04, 5...  
## $ Overtime.Pay (dbl) 37279.96, 0.00, 0.00, 0.00, 0.00, 0.00, 0...  
## $ Other.Pay (dbl) 291.70, 15937.28, 0.00, 0.00, 0.00, 4299....  
## $ Benefits (dbl) 26238.62, 31060.53, 31012.06, 30228.70, 3...  
## $ Total.Pay (dbl) 72327.98, 57309.72, 56923.68, 57503.04, 5...  
## $ Total.Pay...Benefits (dbl) 98566.60, 88370.25, 87935.74, 87731.74, 8...  
## $ Year (dbl) 2014, 2014, 2014, 2014, 2014, 2014, 2014,...  
## $ Notes (fctr) , , , , , , , , , , , , , , , , , , , , ...  
## $ Agency (fctr) State of California, State of California...  
## $ Status (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...

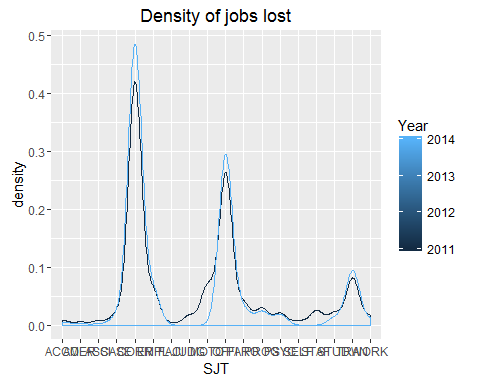
glimpse(JobCutDensity2011)

## Observations: 70,165  
## Variables: 16  
## $ Job\_Title (chr) "ACCOUNTING TECHNICIAN", "ACCOUNTING TECHNI...  
## $ 2011Positions (dbl) 529, 529, 529, 529, 529, 529, 529, 529, 529...  
## $ 2014Positions (dbl) 280, 280, 280, 280, 280, 280, 280, 280, 280...  
## $ SJT (chr) "ACCO", "ACCO", "ACCO", "ACCO", "ACCO", "AC...  
## $ PositionsLost (dbl) 249, 249, 249, 249, 249, 249, 249, 249, 249...  
## $ employee\_name (fctr) ABUTAHA MR SHAHEEN, JOSEFINA M ESTERON, ED...  
## $ base\_pay (dbl) 59711.64, 36877.14, 59711.64, 56975.54, 578...  
## $ overtime\_pay (dbl) 11332.67, 0.00, 0.00, 0.00, 0.00, 0.00, 0.0...  
## $ other\_pay (dbl) 2407.89, 29745.33, 2407.89, 3010.00, 0.00, ...  
## $ total\_benefits (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ total\_pay (dbl) 73452.20, 66622.47, 62119.53, 59985.54, 578...  
## $ total\_pay\_benefits (dbl) 73452.20, 66622.47, 62119.53, 59985.54, 578...  
## $ year (dbl) 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2...  
## $ notes (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ jurisdiction\_name (fctr) State of California, State of California, ...  
## $ Status (chr) "NA", "NA", "NA", "NA", "NA", "NA", "NA", "...

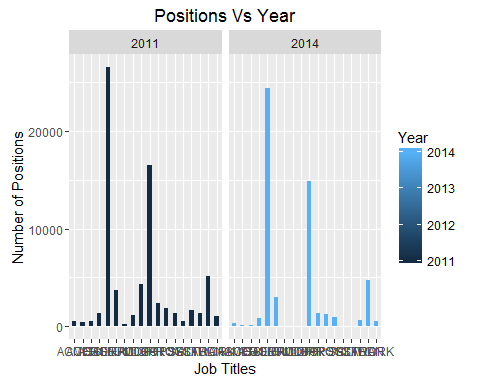
names(JobCutDensity2014)[6]<- paste('Employee\_Name')  
names(JobCutDensity2014)[7]<- paste('Base\_Pay')  
names(JobCutDensity2014)[8]<- paste('Overtime\_Pay')  
names(JobCutDensity2014)[9]<- paste('Other\_Pay')  
names(JobCutDensity2014)[10]<- paste('Benefits')  
names(JobCutDensity2014)[11]<- paste('Total\_Pay')  
names(JobCutDensity2014)[12]<- paste('Total\_Pay\_Benefits')  
names(JobCutDensity2014)[13]<- paste('Year')  
names(JobCutDensity2014)[14]<- paste('Notes')  
names(JobCutDensity2014)[15]<- paste('Agency')  
  
names(JobCutDensity2011)[6]<- paste('Employee\_Name')  
names(JobCutDensity2011)[7]<- paste('Base\_Pay')  
names(JobCutDensity2011)[8]<- paste('Overtime\_Pay')  
names(JobCutDensity2011)[9]<- paste('Other\_Pay')  
names(JobCutDensity2011)[10]<- paste('Benefits')  
names(JobCutDensity2011)[11]<- paste('Total\_Pay')  
names(JobCutDensity2011)[12]<- paste('Total\_Pay\_Benefits')  
names(JobCutDensity2011)[13]<- paste('Year')  
names(JobCutDensity2011)[14]<- paste('Notes')  
names(JobCutDensity2011)[15]<- paste('Agency')  
  
  
JobCutDensity20112014 <- rbind(JobCutDensity2011,JobCutDensity2014)  
  
by(JobCutDensity20112014$Job\_Title,JobCutDensity20112014$Year,summary)

## JobCutDensity20112014$Year: 2011  
## ACCOUNTING TECHNICIAN   
## 529   
## AMERICORPS MEMBERS   
## 363   
## ASSISTANT EXAMINATION PROCTOR   
## 508   
## CASE RECORDS TECHNICIAN   
## 1294   
## CORRECTIONAL COUNSELOR I   
## 1282   
## CORRECTIONAL OFFICER   
## 25290   
## EMPLOYMENT PROGRAM REPRESENTATIVE   
## 3724   
## FACILITY CAPTAIN, CORRECTIONAL INSTITUTION   
## 207   
## JUDGE, SUPERIOR COURT   
## 1080   
## MOTOR VEHICLE FIELD REPRESENTATIVE   
## 3536   
## MOTOR VEHICLE TECHNICIAN   
## 736   
## OFFICE ASSISTANT (GENERAL)   
## 1053   
## OFFICE ASSISTANT (TYPING)   
## 2087   
## OFFICE TECHNICIAN (TYPING)   
## 6878   
## OFFICER, CALIFORNIA HIGHWAY PATROL   
## 6541   
## PAROLE AGENT I ADULT PAROLE   
## 1955   
## PAROLE AGENT II, ADULT PAROLE (SPECIALIST)   
## 357   
## PROGRAM TECHNICIAN   
## 1849   
## PSYCHIATRIC TECHNICIAN   
## 1308   
## SELF-HELP SPONSOR (PART TIME)   
## 504   
## STAFF COUNSEL   
## 535   
## STAFF COUNSEL III (SPECIALIST)   
## 705   
## STAFF ENVIRONMENTAL SCIENTIST   
## 381   
## STUDENT ASSISTANT   
## 1353   
## TRANSPORTATION ENGINEER (CIVIL)   
## 5140   
## WORKERS' COMPENSATION INSURANCE TECHNICIAN   
## 970   
## --------------------------------------------------------   
## JobCutDensity20112014$Year: 2014  
## ACCOUNTING TECHNICIAN   
## 280   
## AMERICORPS MEMBERS   
## 139   
## ASSISTANT EXAMINATION PROCTOR   
## 116   
## CASE RECORDS TECHNICIAN   
## 762   
## CORRECTIONAL COUNSELOR I   
## 1023   
## CORRECTIONAL OFFICER   
## 23442   
## EMPLOYMENT PROGRAM REPRESENTATIVE   
## 3006   
## FACILITY CAPTAIN, CORRECTIONAL INSTITUTION   
## 1   
## JUDGE, SUPERIOR COURT   
## 1   
## MOTOR VEHICLE FIELD REPRESENTATIVE   
## 1   
## MOTOR VEHICLE TECHNICIAN   
## 3   
## OFFICE ASSISTANT (GENERAL)   
## 710   
## OFFICE ASSISTANT (TYPING)   
## 1609   
## OFFICE TECHNICIAN (TYPING)   
## 6323   
## OFFICER, CALIFORNIA HIGHWAY PATROL   
## 6281   
## PAROLE AGENT I ADULT PAROLE   
## 1150   
## PAROLE AGENT II, ADULT PAROLE (SPECIALIST)   
## 141   
## PROGRAM TECHNICIAN   
## 1184   
## PSYCHIATRIC TECHNICIAN   
## 922   
## SELF-HELP SPONSOR (PART TIME)   
## 2   
## STAFF COUNSEL   
## 1   
## STAFF COUNSEL III (SPECIALIST)   
## 1   
## STAFF ENVIRONMENTAL SCIENTIST   
## 1   
## STUDENT ASSISTANT   
## 647   
## TRANSPORTATION ENGINEER (CIVIL)   
## 4759   
## WORKERS' COMPENSATION INSURANCE TECHNICIAN   
## 465

ggplot(JobCutDensity20112014, aes(x=SJT)) +   
 geom\_density(aes(group=Year, colour=Year))+  
 ggtitle("Density of jobs lost")



ggplot(JobCutDensity20112014,aes(x=SJT,fill = Year))+  
 stat\_count(width = 0.5) +  
 facet\_wrap(~Year)+  
 ggtitle("Positions Vs Year")+  
 xlab("Job Titles")+  
 ylab("Number of Positions")+  
 labs(fill="Year")



#Density of Jobs Increased in 2014  
  
  
MajorjobsIncreasedfrom2011 <- join20112014 %>% filter(`2014Positions` - `2011Positions`>200)  
MajorjobsIncreasedfrom2011$SJT <- strtrim(MajorjobsIncreasedfrom2011$Job\_Title, 4)  
MajorjobsIncreasedfrom2011$PositionsIncreased <- MajorjobsIncreasedfrom2011$`2014Positions` - MajorjobsIncreasedfrom2011$`2011Positions`  
  
  
names(Salary2011)[2]<- paste('Job\_Title')  
names(Salary2014)[2]<- paste('Job\_Title')  
  
JobIncreasedDensity2011 <- left\_join(MajorjobsIncreasedfrom2011,Salary2011,"Job\_Title")  
JobIncreasedDensity2011$Status <- 'NA'  
#Ensure Year variable has value 2011  
JobIncreasedDensity2011$year <- 2011  
  
JobIncreasedDensity2014 <- left\_join(MajorjobsIncreasedfrom2011,Salary2014,"Job\_Title")  
#Ensure Year variable has value 2014  
JobIncreasedDensity2014$Year <- 2014  
  
glimpse(JobIncreasedDensity2011)

## Observations: 25,280  
## Variables: 16  
## $ Job\_Title (chr) "ACCOUNTANT TRAINEE", "ACCOUNTANT TRAINEE",...  
## $ 2011Positions (dbl) 255, 255, 255, 255, 255, 255, 255, 255, 255...  
## $ 2014Positions (dbl) 524, 524, 524, 524, 524, 524, 524, 524, 524...  
## $ SJT (chr) "ACCO", "ACCO", "ACCO", "ACCO", "ACCO", "AC...  
## $ PositionsIncreased (dbl) 269, 269, 269, 269, 269, 269, 269, 269, 269...  
## $ employee\_name (fctr) AMY N HA, MAO HER, SUSAN P REYNON, WINNIE ...  
## $ base\_pay (dbl) 38168.01, 38168.01, 38631.57, 38631.57, 381...  
## $ overtime\_pay (dbl) 14376.62, 11255.65, 10081.56, 9133.06, 8967...  
## $ other\_pay (dbl) 1940.24, 1797.10, 1839.78, 1775.50, 1612.77...  
## $ total\_benefits (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ total\_pay (dbl) 54484.87, 51220.76, 50552.91, 49540.13, 487...  
## $ total\_pay\_benefits (dbl) 54484.87, 51220.76, 50552.91, 49540.13, 487...  
## $ year (dbl) 2011, 2011, 2011, 2011, 2011, 2011, 2011, 2...  
## $ notes (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...  
## $ jurisdiction\_name (fctr) State of California, State of California, ...  
## $ Status (chr) "NA", "NA", "NA", "NA", "NA", "NA", "NA", "...

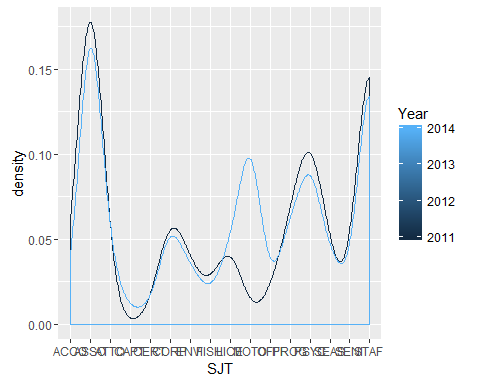
glimpse(JobIncreasedDensity2014)

## Observations: 39,058  
## Variables: 16  
## $ Job\_Title (chr) "ACCOUNTANT TRAINEE", "ACCOUNTANT TRAINEE...  
## $ 2011Positions (dbl) 255, 255, 255, 255, 255, 255, 255, 255, 2...  
## $ 2014Positions (dbl) 524, 524, 524, 524, 524, 524, 524, 524, 5...  
## $ SJT (chr) "ACCO", "ACCO", "ACCO", "ACCO", "ACCO", "...  
## $ PositionsIncreased (dbl) 269, 269, 269, 269, 269, 269, 269, 269, 2...  
## $ Employee.Name (fctr) Eddie D Emory, Maria Delacruz, Maria Cla...  
## $ Base.Pay (dbl) 65406.35, 41446.68, 41594.72, 45854.92, 4...  
## $ Overtime.Pay (dbl) 7792.05, 16218.05, 11164.72, 12496.01, 14...  
## $ Other.Pay (dbl) 84.00, 2047.33, 1431.47, 1601.56, 1046.75...  
## $ Benefits (dbl) 14252.82, 27806.82, 27048.54, 17240.96, 1...  
## $ Total.Pay (dbl) 73282.40, 59712.06, 54190.91, 59952.49, 5...  
## $ Total.Pay...Benefits (dbl) 87535.22, 87518.88, 81239.45, 77193.45, 7...  
## $ Year (dbl) 2014, 2014, 2014, 2014, 2014, 2014, 2014,...  
## $ Notes (fctr) Employee holds multiple positions., , , ...  
## $ Agency (fctr) State of California, State of California...  
## $ Status (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...

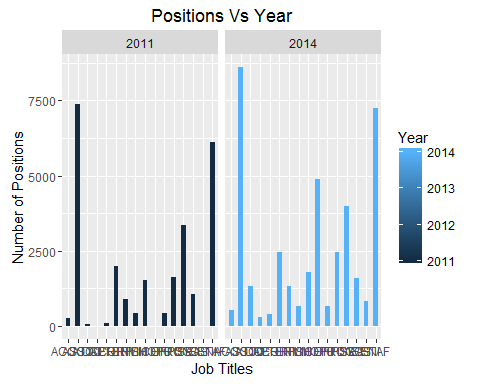
names(JobIncreasedDensity2014)[6]<- paste('Employee\_Name')  
names(JobIncreasedDensity2014)[7]<- paste('Base\_Pay')  
names(JobIncreasedDensity2014)[8]<- paste('Overtime\_Pay')  
names(JobIncreasedDensity2014)[9]<- paste('Other\_Pay')  
names(JobIncreasedDensity2014)[10]<- paste('Benefits')  
names(JobIncreasedDensity2014)[11]<- paste('Total\_Pay')  
names(JobIncreasedDensity2014)[12]<- paste('Total\_Pay\_Benefits')  
names(JobIncreasedDensity2014)[13]<- paste('Year')  
names(JobIncreasedDensity2014)[14]<- paste('Notes')  
names(JobIncreasedDensity2014)[15]<- paste('Agency')  
  
names(JobIncreasedDensity2011)[6]<- paste('Employee\_Name')  
names(JobIncreasedDensity2011)[7]<- paste('Base\_Pay')  
names(JobIncreasedDensity2011)[8]<- paste('Overtime\_Pay')  
names(JobIncreasedDensity2011)[9]<- paste('Other\_Pay')  
names(JobIncreasedDensity2011)[10]<- paste('Benefits')  
names(JobIncreasedDensity2011)[11]<- paste('Total\_Pay')  
names(JobIncreasedDensity2011)[12]<- paste('Total\_Pay\_Benefits')  
names(JobIncreasedDensity2011)[13]<- paste('Year')  
names(JobIncreasedDensity2011)[14]<- paste('Notes')  
names(JobIncreasedDensity2011)[15]<- paste('Agency')  
  
  
JobIncreasedDensity20112014 <- rbind(JobIncreasedDensity2011,JobIncreasedDensity2014)  
  
by(JobIncreasedDensity20112014$Job\_Title,JobIncreasedDensity20112014$Year,summary)

## JobIncreasedDensity20112014$Year: 2011  
## ACCOUNTANT TRAINEE   
## 255   
## ASSOCIATE GOVERNMENTAL PROGRAM ANALYST   
## 7398   
## ATTORNEY   
## 48   
## ATTORNEY III   
## 1   
## CAPTAIN (ADULT INSTITUTION)   
## 1   
## CERTIFIED NURSING ASSISTANT, CORRECTIONAL FACILITY   
## 99   
## CORPSMEMBER, CCC   
## 2003   
## ENVIRONMENTAL SCIENTIST   
## 908   
## FISH AND WILDLIFE SCIENTIFIC AID   
## 432   
## LICENSED VOCATIONAL NURSE, CALIFORNIA DEPARTMENT OF CORRECTIONS AND REHABILITATION   
## 1518   
## MOTOR VEHICLE REPRESENTATIVE   
## 1   
## OFFICE TECHNICIAN (GENERAL)   
## 436   
## PROGRAM TECHNICIAN II   
## 1641   
## PSYCHIATRIC TECHNICIAN (SAFETY)   
## 3367   
## SEASONAL CLERK   
## 1053   
## SENIOR ENVIRONMENTAL SCIENTIST (SUPERVISORY)   
## 1   
## SENIOR ENVIRONMENTAL SCIENTIST(SPECIALIST)   
## 1   
## STAFF SERVICES ANALYST (GENERAL)   
## 3744   
## STAFF SERVICES MANAGER I   
## 2373   
## --------------------------------------------------------   
## JobIncreasedDensity20112014$Year: 2014  
## ACCOUNTANT TRAINEE   
## 524   
## ASSOCIATE GOVERNMENTAL PROGRAM ANALYST   
## 8608   
## ATTORNEY   
## 585   
## ATTORNEY III   
## 756   
## CAPTAIN (ADULT INSTITUTION)   
## 295   
## CERTIFIED NURSING ASSISTANT, CORRECTIONAL FACILITY   
## 385   
## CORPSMEMBER, CCC   
## 2472   
## ENVIRONMENTAL SCIENTIST   
## 1335   
## FISH AND WILDLIFE SCIENTIFIC AID   
## 668   
## LICENSED VOCATIONAL NURSE, CALIFORNIA DEPARTMENT OF CORRECTIONS AND REHABILITATION   
## 1806   
## MOTOR VEHICLE REPRESENTATIVE   
## 4873   
## OFFICE TECHNICIAN (GENERAL)   
## 647   
## PROGRAM TECHNICIAN II   
## 2445   
## PSYCHIATRIC TECHNICIAN (SAFETY)   
## 3994   
## SEASONAL CLERK   
## 1610   
## SENIOR ENVIRONMENTAL SCIENTIST (SUPERVISORY)   
## 295   
## SENIOR ENVIRONMENTAL SCIENTIST(SPECIALIST)   
## 524   
## STAFF SERVICES ANALYST (GENERAL)   
## 4318   
## STAFF SERVICES MANAGER I   
## 2918

ggplot(JobIncreasedDensity20112014, aes(x=SJT)) +   
 geom\_density(aes(group=Year, colour=Year))



ggplot(JobIncreasedDensity20112014,aes(x=SJT,fill = Year))+  
 stat\_count(width = 0.5) +  
 facet\_wrap(~Year)+  
 ggtitle("Positions Vs Year")+  
 xlab("Job Titles")+  
 ylab("Number of Positions")+  
 labs(fill="Year")



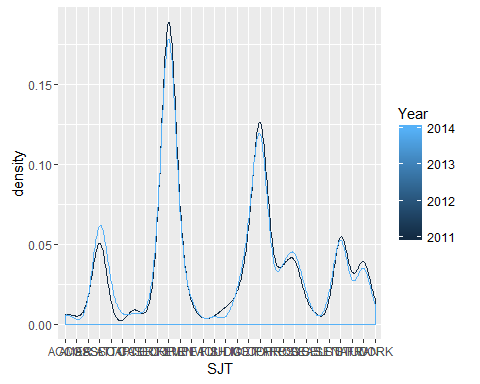
#Job Density by job increased and reduced.  
  
names(JobIncreasedDensity20112014)[5] <- paste('Difference\_Position')  
names(JobCutDensity20112014)[5] <- paste('Difference\_Position')  
  
glimpse(JobIncreasedDensity20112014)

## Observations: 64,338  
## Variables: 16  
## $ Job\_Title (chr) "ACCOUNTANT TRAINEE", "ACCOUNTANT TRAINEE"...  
## $ 2011Positions (dbl) 255, 255, 255, 255, 255, 255, 255, 255, 25...  
## $ 2014Positions (dbl) 524, 524, 524, 524, 524, 524, 524, 524, 52...  
## $ SJT (chr) "ACCO", "ACCO", "ACCO", "ACCO", "ACCO", "A...  
## $ Difference\_Position (dbl) 269, 269, 269, 269, 269, 269, 269, 269, 26...  
## $ Employee\_Name (fctr) AMY N HA, MAO HER, SUSAN P REYNON, WINNIE...  
## $ Base\_Pay (dbl) 38168.01, 38168.01, 38631.57, 38631.57, 38...  
## $ Overtime\_Pay (dbl) 14376.62, 11255.65, 10081.56, 9133.06, 896...  
## $ Other\_Pay (dbl) 1940.24, 1797.10, 1839.78, 1775.50, 1612.7...  
## $ Benefits (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ Total\_Pay (dbl) 54484.87, 51220.76, 50552.91, 49540.13, 48...  
## $ Total\_Pay\_Benefits (dbl) 54484.87, 51220.76, 50552.91, 49540.13, 48...  
## $ Year (dbl) 2011, 2011, 2011, 2011, 2011, 2011, 2011, ...  
## $ Notes (chr) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ Agency (fctr) State of California, State of California,...  
## $ Status (chr) "NA", "NA", "NA", "NA", "NA", "NA", "NA", ...

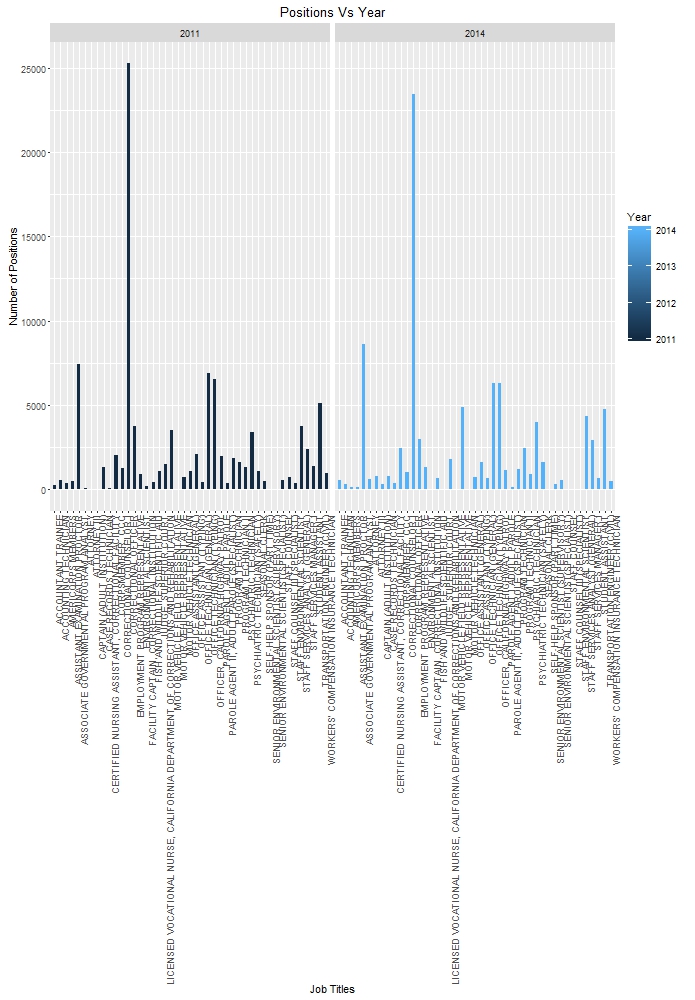
glimpse(JobCutDensity20112014)

## Observations: 123,135  
## Variables: 16  
## $ Job\_Title (chr) "ACCOUNTING TECHNICIAN", "ACCOUNTING TECHN...  
## $ 2011Positions (dbl) 529, 529, 529, 529, 529, 529, 529, 529, 52...  
## $ 2014Positions (dbl) 280, 280, 280, 280, 280, 280, 280, 280, 28...  
## $ SJT (chr) "ACCO", "ACCO", "ACCO", "ACCO", "ACCO", "A...  
## $ Difference\_Position (dbl) 249, 249, 249, 249, 249, 249, 249, 249, 24...  
## $ Employee\_Name (fctr) ABUTAHA MR SHAHEEN, JOSEFINA M ESTERON, E...  
## $ Base\_Pay (dbl) 59711.64, 36877.14, 59711.64, 56975.54, 57...  
## $ Overtime\_Pay (dbl) 11332.67, 0.00, 0.00, 0.00, 0.00, 0.00, 0....  
## $ Other\_Pay (dbl) 2407.89, 29745.33, 2407.89, 3010.00, 0.00,...  
## $ Benefits (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ Total\_Pay (dbl) 73452.20, 66622.47, 62119.53, 59985.54, 57...  
## $ Total\_Pay\_Benefits (dbl) 73452.20, 66622.47, 62119.53, 59985.54, 57...  
## $ Year (dbl) 2011, 2011, 2011, 2011, 2011, 2011, 2011, ...  
## $ Notes (chr) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...  
## $ Agency (fctr) State of California, State of California,...  
## $ Status (chr) "NA", "NA", "NA", "NA", "NA", "NA", "NA", ...

JobDensity <- rbind(JobIncreasedDensity20112014,JobCutDensity20112014)  
JobDensity$Difference\_Position <- (JobDensity$`2014Positions`-JobDensity$`2011Positions`)  
  
#Density of number of positions per title VS Year  
ggplot(JobDensity, aes(x=SJT)) +   
 geom\_density(aes(group=Year, colour=Year))



#Positions per Job Title grouped by year  
job20112014Plot <- ggplot(JobDensity,aes(x=Job\_Title,fill = Year))+  
 stat\_count(width = 0.5) +  
 facet\_wrap(~Year)+  
 ggtitle("Positions Vs Year")+  
 xlab("Job Titles")+  
 ylab("Number of Positions")+  
 labs(fill="Year")  
  
job20112014Plot+ theme(axis.text.x = element\_text(angle = 90, hjust = 1))



JobDensity$Base\_Pay[is.na(JobDensity$Base\_Pay)] <- 0  
  
JobMeanPay <- JobDensity %>% group\_by(factor(Job\_Title),Difference\_Position) %>%  
 summarise(MeanPay=mean(Base\_Pay))  
  
names(JobMeanPay)[1] <- paste('Job\_Title')  
  
  
  
glimpse(JobMeanPay)

## Observations: 45  
## Variables: 3  
## $ Job\_Title (fctr) ACCOUNTANT TRAINEE, ACCOUNTING TECHNICIAN...  
## $ Difference\_Position (dbl) 269, -249, -224, -392, 1210, 537, 756, 295...  
## $ MeanPay (dbl) 32171.2493, 32753.9444, 6612.8010, 727.277...

#Jobs lost and added in 2014(cpmpared to jobs in 2011)  
#scale\_x\_log10()  
  
ggplot(data=JobMeanPay, aes(x=MeanPay, y=Difference\_Position)) +  
 geom\_bar(stat="identity",width = .1)+  
 scale\_y\_continuous(limits = c(-3000, 4000), breaks = seq(-3000, 4000, 10)) +  
 scale\_x\_continuous() +  
 scale\_x\_log10()+  
 ylim(c(-3000,4000))+  
 ylab('<--Positions Lost from 2011 . Positions Added in 2014 -->')+  
 xlab('Mean Pay')

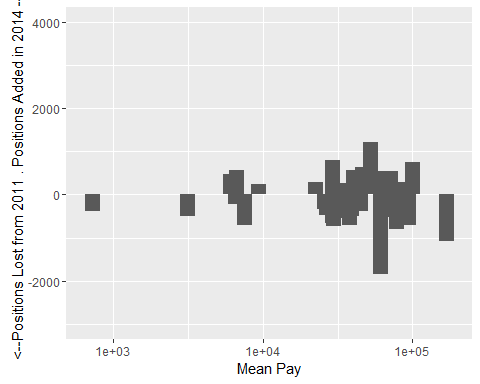
## Scale for 'x' is already present. Adding another scale for 'x', which  
## will replace the existing scale.

## Scale for 'y' is already present. Adding another scale for 'y', which  
## will replace the existing scale.

## Warning: Removed 2 rows containing missing values (position\_stack).

## Warning: Stacking not well defined when ymin != 0

## Warning: position\_stack requires non-overlapping x intervals



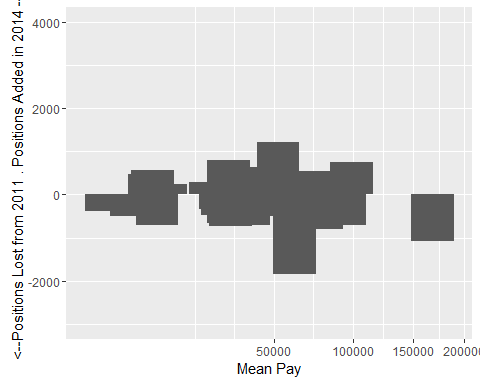
#scale\_x\_sqrt()  
ggplot(data=JobMeanPay, aes(x=MeanPay, y=Difference\_Position)) +  
 geom\_bar(stat="identity",width = 50)+  
 scale\_y\_continuous(limits = c(-3000, 4000), breaks = seq(-3000, 4000, 10)) +  
 scale\_x\_sqrt()+  
 ylim(c(-3000,4000))+  
 ylab('<--Positions Lost from 2011 . Positions Added in 2014 -->')+  
 xlab('Mean Pay')

## Scale for 'y' is already present. Adding another scale for 'y', which  
## will replace the existing scale.

## Warning: Removed 2 rows containing missing values (position\_stack).

## Warning: Stacking not well defined when ymin != 0

## Warning: position\_stack requires non-overlapping x intervals



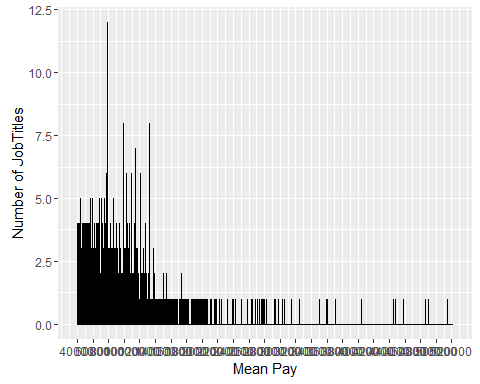
#mean salaries of jobs for 2011 and 2014  
  
MeanPay2011 <- Salary2011 %>% group\_by(factor(Job\_Title)) %>%  
 summarise(mean=mean(total\_pay), sd=sd(total\_pay))  
#MeanPay2011 <- Salary2011 %>% group\_by(factor(job\_title)) %>%  
#summarise(mean=mean(total\_pay))  
#MeanPay2011$Year <- '2011'  
  
names(MeanPay2011)[1] <- paste("Job\_Title")  
names(MeanPay2011)[2] <- paste("2011Mean")  
names(MeanPay2011)[3] <- paste("2011SD")  
  
  
  
MeanPay2014 <- Salary2014 %>% group\_by(factor(Job\_Title)) %>%  
 summarise(mean=mean(Total.Pay), sd=sd(Total.Pay))  
  
#MeanPay2014 <- Salary2014 %>% group\_by(factor(Job.Title)) %>%  
# summarise(mean=mean(Total.Pay))  
  
names(MeanPay2014)[1] <- paste("Job\_Title")  
names(MeanPay2014)[2] <- paste("2014Mean")  
names(MeanPay2014)[3] <- paste("2014SD")  
  
  
MeanPay2011 <- left\_join(Jobtitle2011byCount,MeanPay2011,"Job\_Title")  
MeanPay2014 <- left\_join(Jobtitle2014byCount,MeanPay2014,"Job\_Title")  
  
names(MeanPay2011)[1] <- paste("Job\_Title")  
names(MeanPay2011)[2] <- paste("2011Positions")  
names(MeanPay2011)[3] <- paste("2011Mean")  
names(MeanPay2011)[4] <- paste("2011SD")  
names(MeanPay2014)[1] <- paste("Job\_Title")  
names(MeanPay2014)[2] <- paste("2014Positions")  
names(MeanPay2014)[3] <- paste("2014Mean")  
names(MeanPay2014)[4] <- paste("2014SD")  
  
JobsandMeanPay20112014 <- full\_join(MeanPay2011,MeanPay2014,"Job\_Title")

## Warning in outer\_join\_impl(x, y, by$x, by$y): joining factors with  
## different levels, coercing to character vector

#update NA to 0 for all variables.  
  
JobsandMeanPay20112014$`2011SD`[is.na(JobsandMeanPay20112014$`2011SD`)] <- 0  
JobsandMeanPay20112014$`2014Positions`[is.na(JobsandMeanPay20112014$`2014Positions`)] <- 0  
JobsandMeanPay20112014$`2014Mean`[is.na(JobsandMeanPay20112014$`2014Mean`)] <- 0  
JobsandMeanPay20112014$`2014SD`[is.na(JobsandMeanPay20112014$`2014SD`)] <- 0  
  
#More Jobs with less mean pay   
ggplot(aes(x = MeanPay2011$`2011Mean`), data = MeanPay2011) +  
 geom\_freqpoly( binwidth=20) +   
 scale\_x\_continuous(limits = c(40000, 522594), breaks = seq(40000, 522594, 20000)) +   
 xlab('Mean Pay') +   
 ylab('Number of JobTitles')

## Warning: Removed 748 rows containing non-finite values (stat\_bin).

## Warning: Removed 2 rows containing missing values (geom\_path).



#Higer the basepay, lesser the positions  
ggplot(data=MeanPay2011, aes(x=MeanPay2011$`2011Mean`, y=MeanPay2011$`2011Positions`)) +  
 geom\_bar(stat="identity")+  
 scale\_x\_continuous(limits = c(400000, 522594), breaks = seq(400000, 522594, 10000)) +  
 ylim(c(0,25))+  
 ylab('Number of positions')+  
 xlab('Mean Pay')

## Warning: Removed 4039 rows containing missing values (position\_stack).

