PS2

Mashhood Syed

October 17, 2014

This dataset named "ts_data" is a set of observations of tech startup companies that have received VC funding. The dataset is rich with categorical (industry sector, funding status) variables, monetary variables, time variables, and location variables. This dataset was found at crunchbase.com.

We know that the San Francisco Bay Area has a dense concentration of tech startups. But what about other regions of CA? How does Los Angeles's startup scene look for softare companies? In this assignment I take a look at a few different variables to get a better idea of what LA's tech scene tells us.

```
      Summary for the Variable "Category_list"
```

```
setwd("~/Desktop/PS2")
ts_file = "~/Desktop/PS2/crunchbase_monthly_export.csv"
ts_data = read.csv(ts_file, header=T)
#cleaning up the data
funding = ts_data$funding_total_usd
test <- funding
destring <- function(x,keep="0-9.") {
    return( as.numeric(gsub(paste("[^",keep,"]+",sep=""),"",x)) )
}
funding_num = destring(test)
#creating subsets
ts_data = cbind(ts_data, funding_num)
ts_data3 = subset(ts_data, market == " Software ")
ts_data_final = ts_data3

cali = subset(ts_data_final, state_code == "CA")
sfbay = subset(cali, region == "SF Bay Area")
la = subset(cali, region == "Los Angeles")
attach(la)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     funding_num, region
```

```
#summaries
options(scipen=999)
summary(sfbay)
```

```
##                        permalink             name
##  /organization/15five       :  1   15Five       :  1
##  /organization/28msec       :  1   28msec       :  1
##  /organization/3gear-systems:  1   2DOLife.com  :  1
##  /organization/3ten8        :  1   3Gear Systems:  1
##  /organization/4blox        :  1   3TEN8        :  1
##  /organization/ablesky      :  1   4Blox        :  1
##  (Other)                    :668   (Other)      :668
```

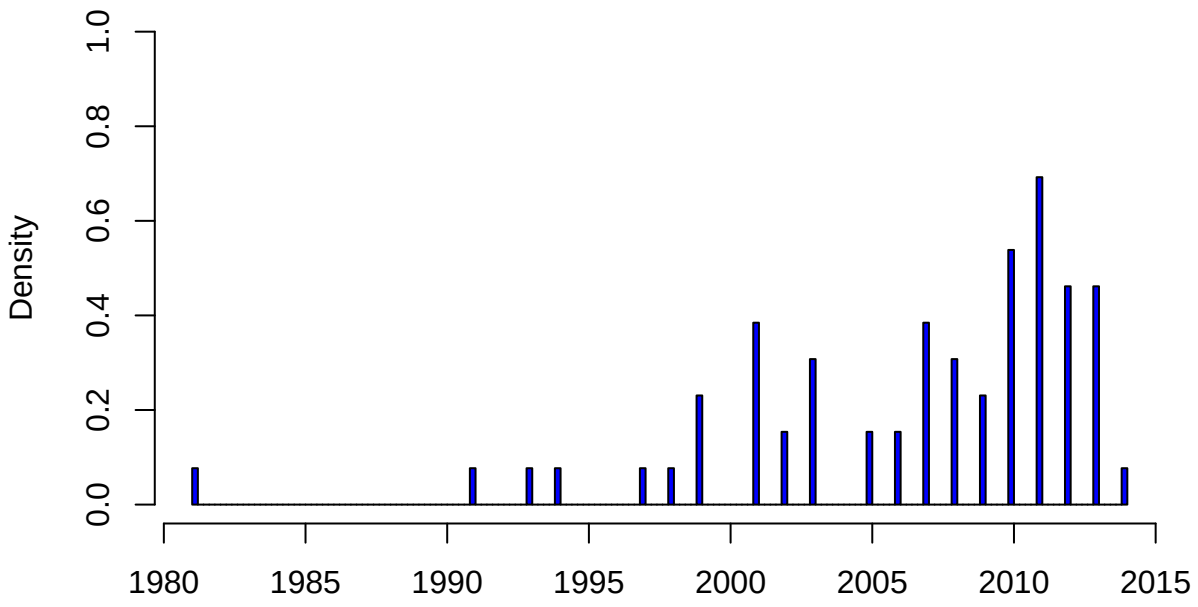
```
##                homepage_url                    category_list
##                      : 40   |Software|               :525
##  http://15five.com   :  1   |Finance|Software|       :  3
##  http://2DOLife.com  :  1   |B2B|SaaS|Software|      :  2
##  http://5apes.com    :  1   |Cloud Security|Software|:  2
##  http://ablesky.com  :  1   |Networking|Software|    :  2
##  http://actionrun.com:  1   |Open Source|Software|   :  2
##  (Other)             :629   (Other)                  :138
##              market     funding_total_usd       status     country_code
##   Software      :674    -        : 66              : 14   USA    :674
##                 :  0   10,000,000: 15     acquired :119          :  0
##   3D            :  0   1,000,000 : 13     closed   : 52   ALB    :  0
##   3D Printing   :  0   2,000,000 : 12     operating:489   ARE    :  0
##   3D Technology :  0   4,000,000 : 10                     ARG    :  0
##   Accounting    :  0   500,000   : 10                     ARM    :  0
##  (Other)        :  0   (Other)   :548                     (Other):  0
##    state_code          region               city     funding_rounds
##  CA     :674   SF Bay Area:674   San Francisco:209   Min.   : 1.00
##         :  0              :  0   Palo Alto    : 63   1st Qu.: 1.00
##  AB     :  0   <c4>vry    :  0   Mountain View: 57   Median : 1.00
##  AK     :  0   <e2>an     :  0   Santa Clara  : 53   Mean   : 1.93
##  AL     :  0   A Coruna   :  0   Sunnyvale    : 43   3rd Qu.: 2.00
##  AR     :  0   Aachen     :  0   San Jose     : 42   Max.   :14.00
##  (Other):  0   (Other)    :  0   (Other)      :207
##    founded_at  founded_month founded_quarter  founded_year
##         : 90          : 90          : 90     Min.   :1979
##  1/1/11 : 37   2011-01: 38   2011-Q1: 44     1st Qu.:2003
##  1/1/12 : 33   2012-01: 33   2012-Q1: 37     Median :2008
##  1/1/03 : 27   2003-01: 28   2010-Q1: 33     Mean   :2006
##  1/1/10 : 26   2010-01: 27   2003-Q1: 28     3rd Qu.:2011
##  1/1/04 : 25   2004-01: 25   2004-Q1: 26     Max.   :2014
##  (Other):436   (Other):433   (Other):416     NA's   :90
##  first_funding_at last_funding_at  funding_num
##  1/1/08 :  8      10/1/12:  6     Min.   :      5500
##  1/1/13 :  7      1/1/13 :  5     1st Qu.:   1500000
##  10/1/12:  6      1/1/08 :  4     Median :   5670000
##  8/1/12 :  6      4/1/13 :  4     Mean   :  19068487
##  1/1/12 :  5      8/1/12 :  4     3rd Qu.:  16000000
##  1/1/05 :  4      1/1/06 :  3     Max.   :1160166511
##  (Other):638      (Other):648     NA's   :66
```

```
summary(la)
```

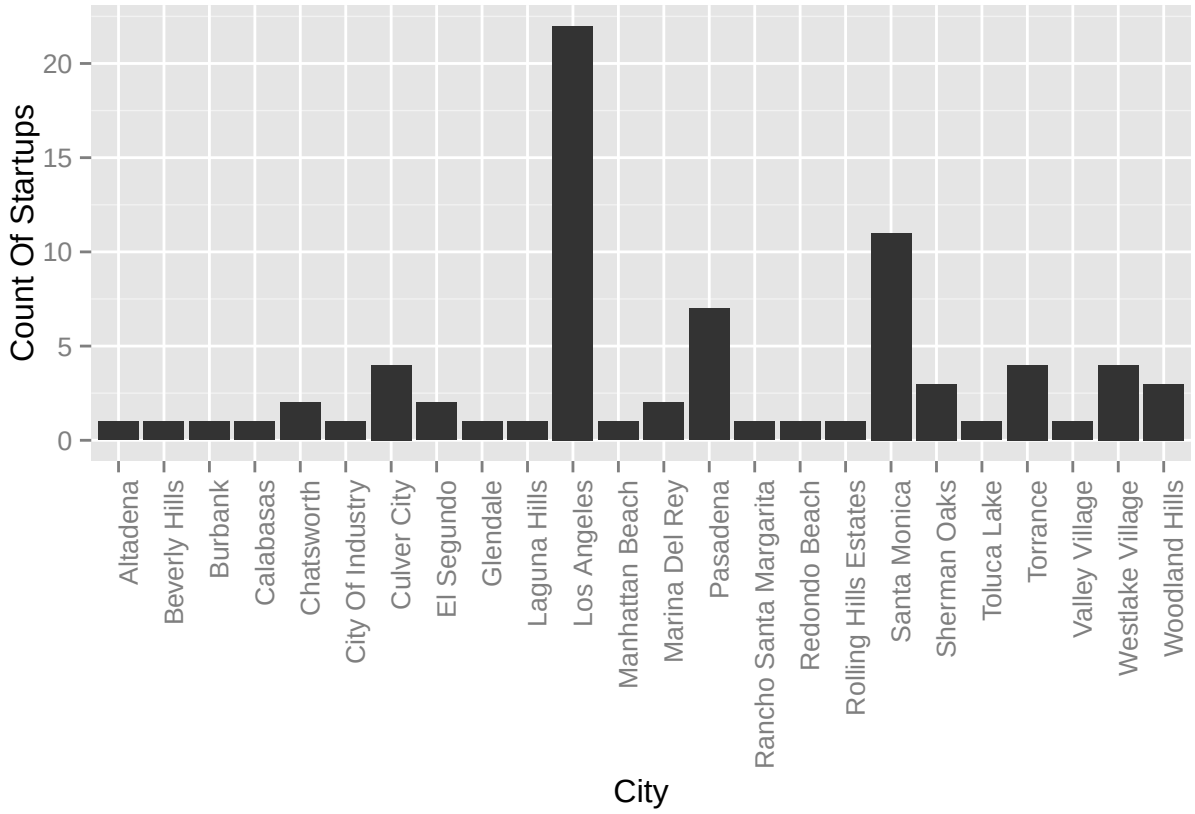
```
##                          permalink                name
##  /organization/5k-fans        : 1   5k Fans        : 1
##  /organization/acid-labs      : 1   Acid Labs      : 1
##  /organization/agency-system  : 1   Agency Systems : 1
##  /organization/anafore        : 1   Anafore        : 1
##  /organization/aspyra         : 1   Aspyra         : 1
##  /organization/awr-corporation: 1   AWR Corporation: 1
##  (Other)                      :71   (Other)        :71
##                  homepage_url
##                       : 6
##  http://accesshsi.com  : 1
```

```
##  http://aspyra.com     : 1
##  http://awrcorp.com    : 1
##  http://borderstylo.com: 1
##  http://brevityv.com   : 1
##  (Other)               :66
##                                                                  category_list
##  |Software|                                                              :58
##  |Artificial Intelligence|Software|                                      : 1
##  |Automotive|Social Commerce|Group Buying|Retail|Discounts|Coupons|Software|: 1
##  |Cloud Data Services|Developer APIs|Software|                          : 1
##  |Data Security|Ediscovery|Software|                                    : 1
##  |Entertainment|Games|Software|                                         : 1
##  (Other)                                                               :14
##              market    funding_total_usd       status    country_code
##   Software      :77    -        :15              : 1   USA    :77
##                 : 0   1,000,000 : 3      acquired :10          : 0
##   3D            : 0   12,000,000: 2      closed   : 8   ALB    : 0
##   3D Printing   : 0   125,000   : 2      operating:58   ARE    : 0
##   3D Technology : 0   2,500,000 : 2                     ARG    : 0
##   Accounting    : 0   900,000   : 2                     ARM    : 0
##  (Other)        : 0   (Other)   :51                     (Other): 0
##    state_code         region                 city    funding_rounds
##  CA     :77   Los Angeles:77   Los Angeles     :22   Min.   :1.00
##         : 0              : 0   Santa Monica    :11   1st Qu.:1.00
##  AB     : 0   <c4>vry    : 0   Pasadena        : 7   Median :1.00
##  AK     : 0   <e2>an     : 0   Culver City     : 4   Mean   :1.53
##  AL     : 0   A Coruna   : 0   Torrance        : 4   3rd Qu.:2.00
##  AR     : 0   Aachen     : 0   Westlake Village: 4   Max.   :4.00
##  (Other): 0   (Other)    : 0   (Other)         :25
##    founded_at founded_month founded_quarter  founded_year
##         :12          :12           :12      Min.   :1981
##  1/1/10 : 5   2001-01: 5    2010-Q1: 6      1st Qu.:2003
##  1/1/11 : 5   2010-01: 5    2001-Q1: 5      Median :2008
##  1/1/01 : 4   2011-01: 5    2011-Q1: 5      Mean   :2006
##  1/1/13 : 4   2012-01: 4    2012-Q1: 5      3rd Qu.:2011
##  1/1/03 : 3   2013-01: 4    2013-Q1: 5      Max.   :2014
##  (Other):44   (Other):42    (Other):39      NA's   :12
##  first_funding_at last_funding_at  funding_num
##  8/14/14: 2       8/14/14: 2      Min.   :   12000
##  9/21/12: 2       1/1/10 : 1      1st Qu.:  752722
##  1/1/10 : 1       1/1/13 : 1      Median : 2500000
##  1/1/12 : 1       1/11/12: 1      Mean   : 6746043
##  1/1/13 : 1       1/14/14: 1      3rd Qu.: 6350000
##  1/11/12: 1       1/28/13: 1      Max.   :90000000
##  (Other):69       (Other):70      NA's   :15
```

```
library(ggplot2)
hist(la$founded_year, prob=T, ylim=c(0, 1), breaks = 150, main = " ", xlab = " ", col=4)
```



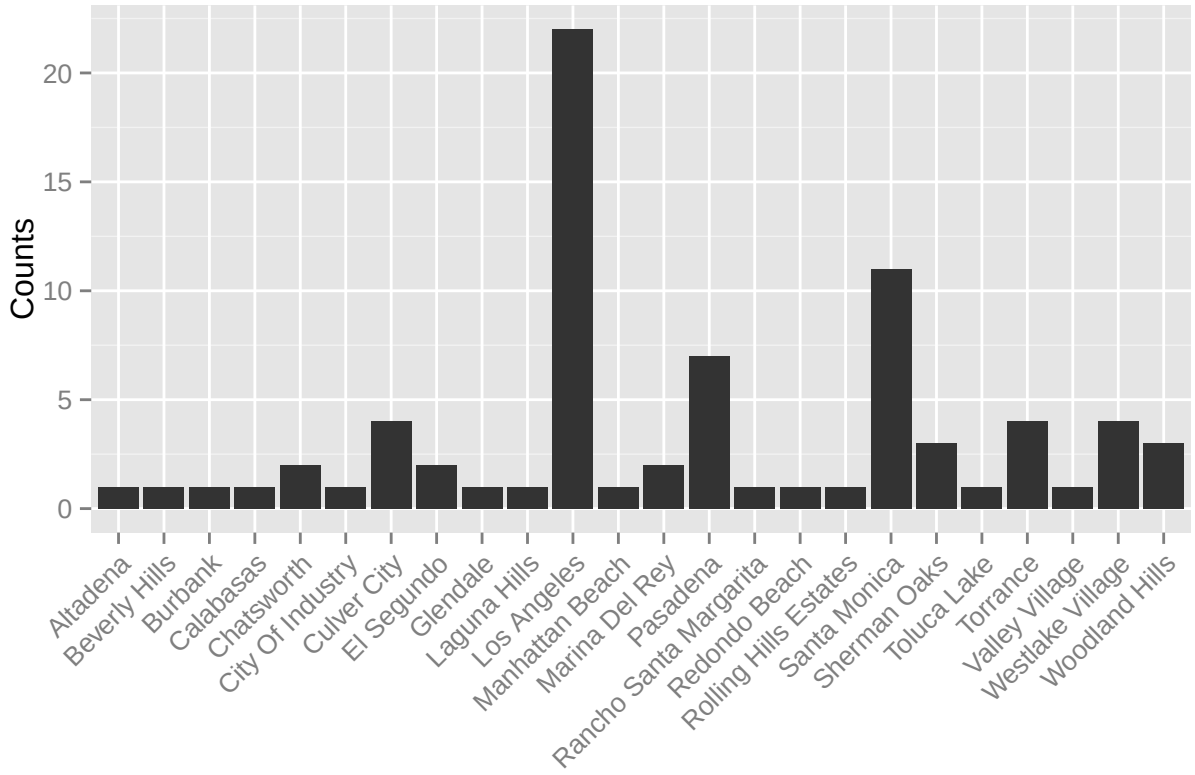

```
ggplot(la, aes(x=city)) + geom_histogram(aes(y=..count..)) + ylab( "Count Of Startups") + xlab( "City")
```



Compare Qualitative vs. Qualitative

```
Comparing "city" vs. "funding amount"
```

```
ggplot(la, aes(x=city)) + geom_histogram( aes(y=..count..)) + ylab( "Counts") + xlab( " ") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

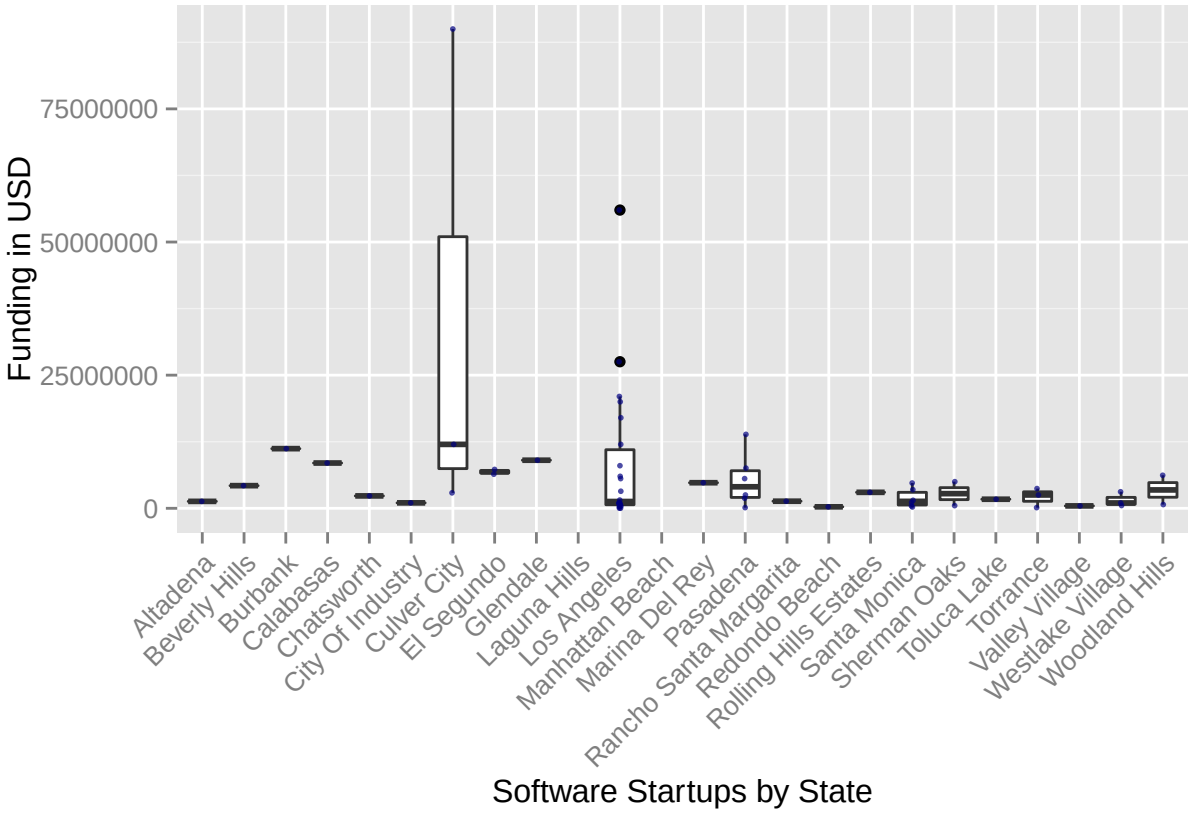


Comparing Qualitative vs. Quantitative

```
    Box Plot of Funding by City
```

```
ggplot(la, aes(x=city, y=funding_num)) + geom_boxplot(width=0.90) + ylab( "Funding in USD") + xlab( "So
```

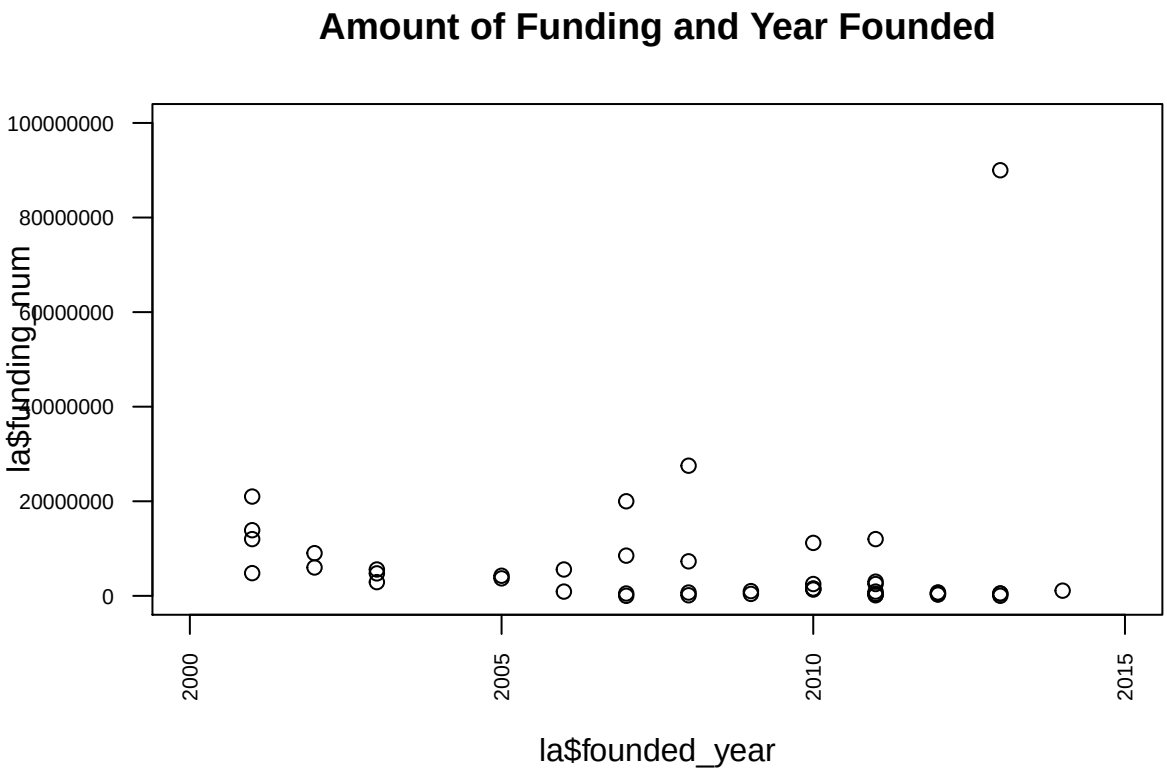
```
## Warning: Removed 15 rows containing non-finite values (stat_boxplot).
## Warning: Removed 15 rows containing missing values (geom_point).
```



Comparing Quantitative vs. Quantitative

Plot of Founded Year and Funding Amount

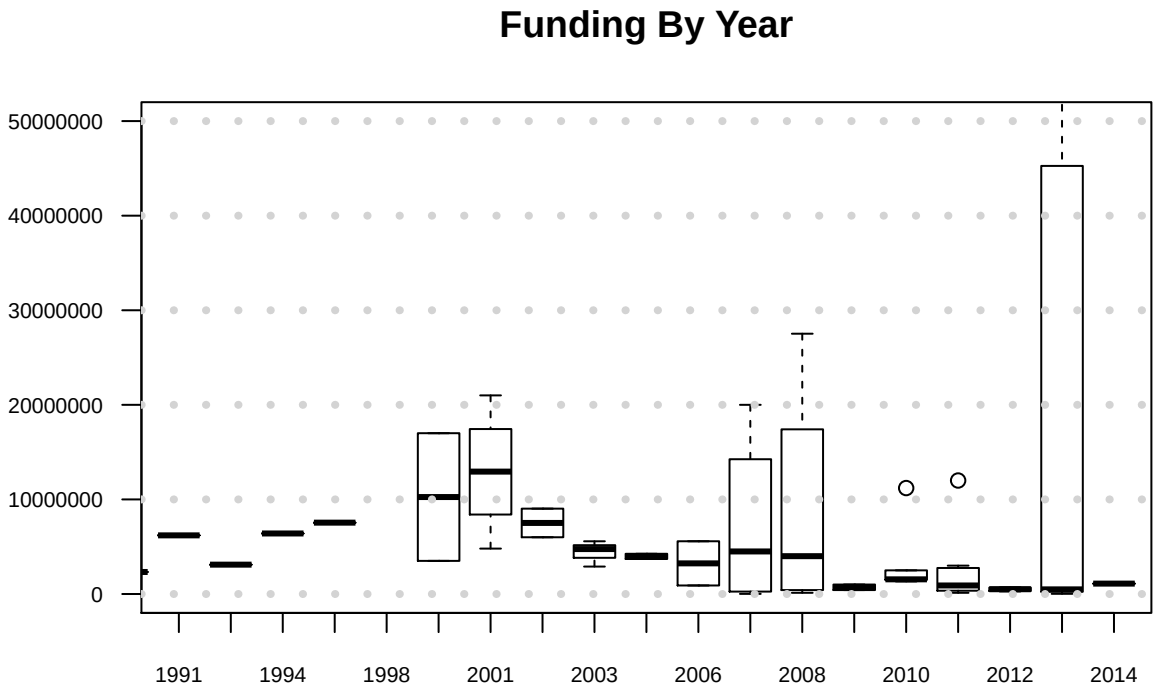
```
plot(la$funding_num~la$founded_year, ylim=c(0, 100000000), xlim=c(2000, 2015), las=2, cex.axis=.65, mai
```



Comparing Quantitative vs. Quantitative

Boxplot of The Amount of Funding By Year

```
boxplot(la$funding_num~la$founded_year, las=1, cex.axis=.65,ylim=c(0, 50000000), main="Funding By Year"
grid(nx=NA, ny=NULL, lwd = 4)
```




```
options(scipen=999)
quantile(sfbay$funding_num, probs=c(0, 0.25, .5, 0.75, 1), na.rm=TRUE)
```

```
##          0%         25%         50%         75%        100%
##        5500     1500000     5670000    16000000 1160166511
```

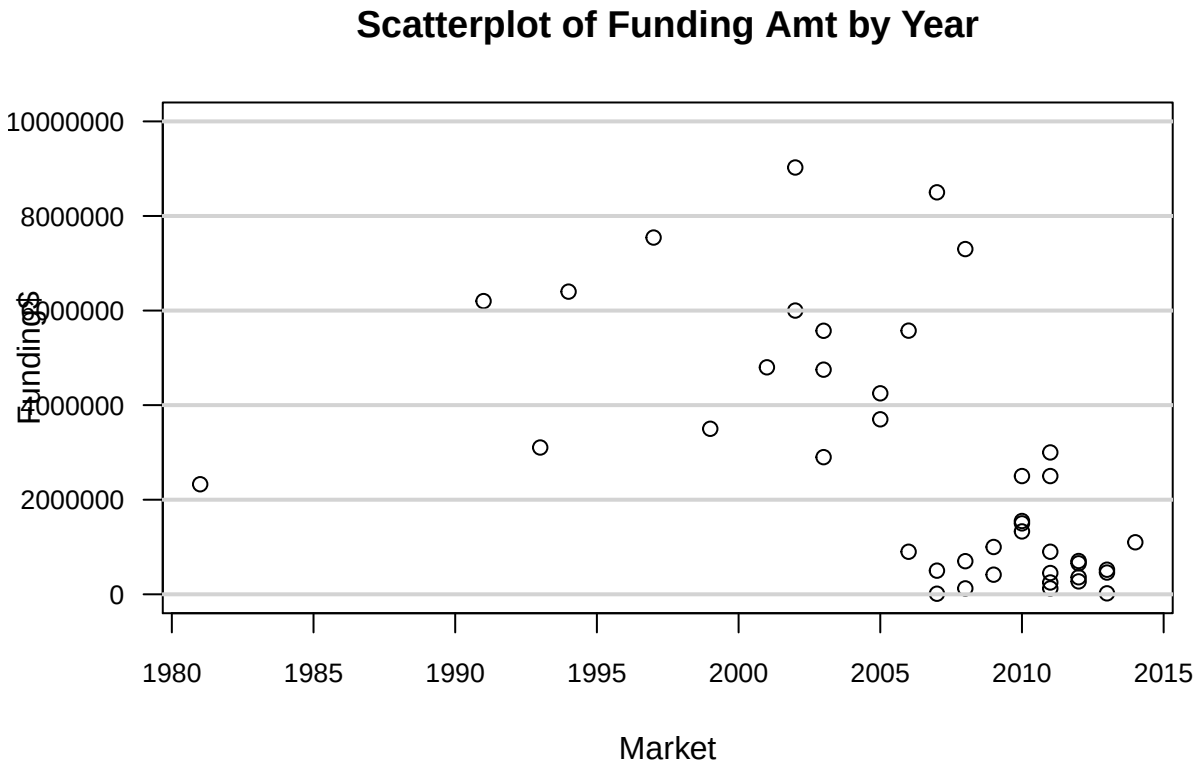
Comparing Quantitative vs. Quantitative

Scatterplot of Funding Amt vs. Year

```
plot(la$founded_year, la$funding_num, main="Scatterplot of Funding Amt by Year", xlab="Market", ylab="Fu
```

```
## NULL
```

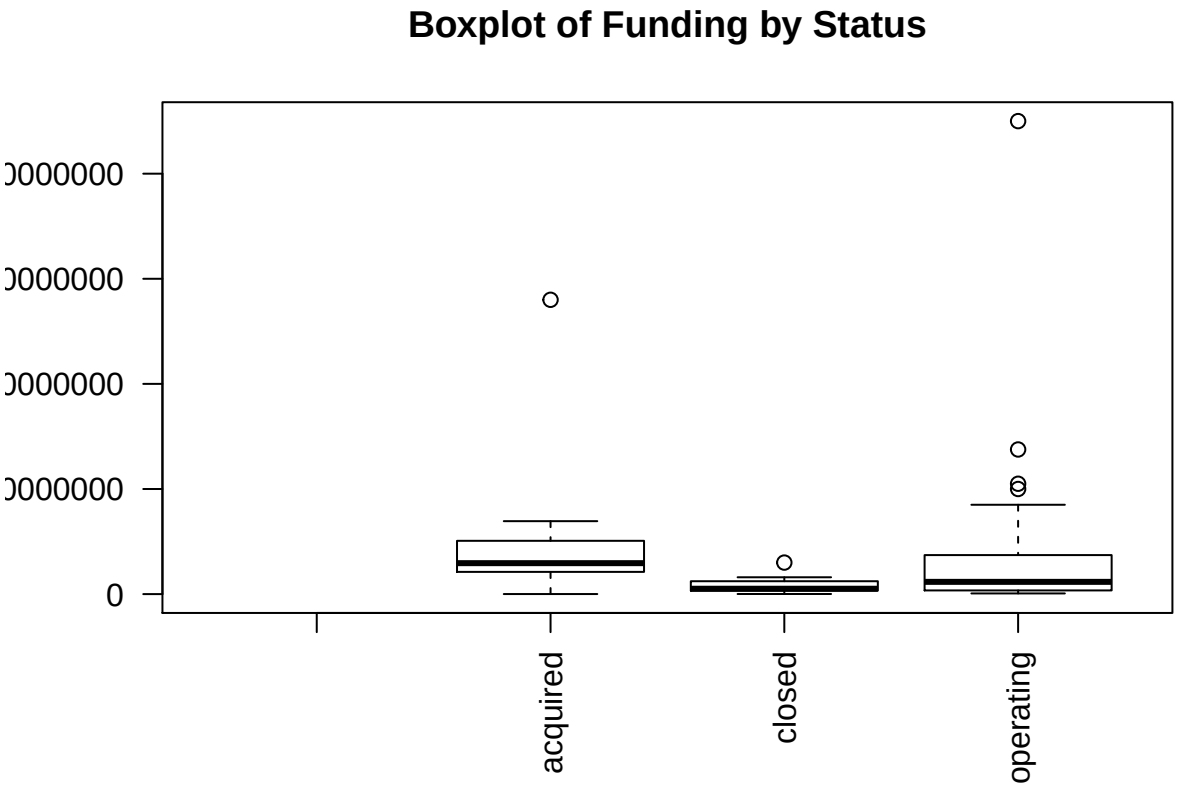
```
grid(nx=NA, ny=NULL, lty=1, lwd = 2)
```




```
quantile(la$funding_num, probs=c(0, 0.25, .5, 0.75, 1), na.rm=TRUE)
```

```
##        0%       25%       50%       75%      100%
##     12000    752722   2500000   6350000 90000000
```

```r
boxplot(la$funding_num ~ la$status, las=2, main="Boxplot of Funding by Status")
```



```
In this scatterplot, we look at the different types of software startups
the year founded and the amount of funding they have received.
```

```
ggplot(la, aes( x=founded_year, y=funding_num, colour = category_list)) + geom_point(alpha=0.8) + stat_
```

```
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 18 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 1 rows containing missing values (stat_smooth).
## Warning: Removed 25 rows containing missing values (geom_point).
```

Automotive|Social Commerce|Group Buying|Retail|Discounts|Coupons|Software|

Cloud Data Services|Developer APIs|Software|

Data Security|Ediscovery|Software|

Entertainment|Games|Software|

Entrepreneur|Internet|Sales and Marketing|B2B|Small and Medium Businesses|Domains|Web Developmer

Finance|Social Network Media|Social Media Management|Software|

nternet|Software|

Mac|Computers|iPad|Android|iPhone|Technology|Software|

Mobile Payments|Software|

Music|Consumer Electronics|Mac|Software|

SaaS|Software|

Sales and Marketing|Artists Globally|Music|Software|

Search|Semantic Web|Neuroscience|Natural Language Processing|Software|

Social Media|Communications Infrastructure|Software|

Software|

Software|E−Commerce|Mobile|Web Hosting|

Software|SEO|Internet Marketing|Web Design|Freelancers|Web Development|

Television|Software|

Video|Messaging|Entertainment|Software|