

# HW #4

*Siavash Mortezaei, Jason Liu, Holly Capell, Sangyu Shen, Kunal Kotian*

*September 13, 2017*

---

## 4A: Aggregate Functions and Dates

**Q. (1)**

(a)

```
qn_1a <- c("select count(1) from stocks2016.d2010;")  
dbGetQuery(first_database, qn_1a)
```

```
##      count  
## 1 1679323
```

(b)

```
qn_1b <- c("select count(1) from (select distinct cusip  
      from stocks2016.d2010) as uniqcusips;")  
dbGetQuery(first_database, qn_1b)
```

```
##      count  
## 1 7108
```

(c)

```
qn_1c <- c("select count(1) from (select distinct cusip  
      from stocks2016.d2011) as uniqcusips;")  
dbGetQuery(first_database, qn_1c)
```

```
##      count  
## 1 7133
```

(d)

```
qn_1d <- c("select cusip from  
(select cusip, count(1) as ct  
      from stocks2016.d2010 group by cusip) as grouped  
where ct < 50;")  
dbGetQuery(first_database, qn_1d)
```

```
##      cusip  
## 1 85815M10  
## 2 78381P10  
## 3 Y1771G10  
## 4 00768Y50  
## 5 G7997810  
## 6 71340910  
## 7 92204267  
## 8 81663A10  
## 9 19248Y10
```

```

## 10 74727D30
## [ reached getOption("max.print") -- omitted 183 rows ]
(e)
qn_1e <- c("select count(1) from
(select cusip from
(select cusip, count(1) as ct
   from stocks2016.d2010 group by cusip) as grouped
where ct < 50) as smallcusips;")
dbGetQuery(first_database,qn_1e)

##   count
## 1    193

(f)
qn_1f <- c("select
sum(case when cusips_count < 50 then 1 else 0 end) as less_than_50,
sum(case when cusips_count > 100 then 1 else 0 end) as more_than_100
from (
select count(cusip) as cusips_count from stocks2016.d2010
group by cusip) q1;")
dbGetQuery(first_database,qn_1f)

##   less_than_50 more_than_100
## 1          193           6754

(g)
qn_1g <- c("select numtype, count(1)
from
(select case
      when cusips_count < 50 then 'less than 50'
      when cusips_count > 100 then 'greater than 100'
      end as numtype
from (
select count(cusip) as cusips_count
from stocks2016.d2010
group by cusip) q1) q2
where numtype is not null
group by numtype;")
dbGetQuery(first_database,qn_1g)

##             numtype count
## 1 greater than 100  6754
## 2 less than 50    193

(h)
qn_1h <- c("select numtype, avg(annual_vol) as avg_annual_vol
from
(select annual_vol, case
      when days_count < 50 then 'less than 50'
      when days_count between 50 and 100 then 'between 50 and 100'
      when days_count > 100 then 'greater than 100'
      end as numtype
from (

```

```

    select cusip, sum(vol) as annual_vol, count(retdate) as days_count
    from stocks2016.d2010
    group by cusip) as q1) as q2
where numtype is not null
group by numtype;")
dbGetQuery(first_database,qn_1h)

##           numtype avg_annual_vol
## 1 between 50 and 100      33819222
## 2 greater than 100       313563246
## 3 less than 50          20505355

(i)
qn_1i <- c("select numtype, avg(avg_vol) as avg_annual_vol
from
(select avg_vol, case
when days_count < 50 then 'less than 50'
when days_count between 50 and 100 then 'between 50 and 100'
when days_count > 100 then 'greater than 100'
end as numtype
from (
select cusip, avg(vol) as avg_vol, count(retdate) as days_count
from stocks2016.d2010
group by cusip) as q1) as q2
where numtype is not null
group by numtype;")
dbGetQuery(first_database,qn_1i)

##           numtype avg_annual_vol
## 1 between 50 and 100      444935.5
## 2 greater than 100        1260723.1
## 3 less than 50            968185.7

(j)
qn_1j <- c("select count(*)
from (select distinct permno
      from (select permno from stocks2016.d2010
            where vol*prc > 100000000 )q1)q2;")
dbGetQuery(first_database,qn_1j)

##   count
## 1 1643

(k)
qn_1k <- c("select count(distinct permno)::float/(select count(distinct permno)
from stocks2016.d2010)
from stocks2016.d2010 where prc*vol > 100000000")
dbGetQuery(first_database,qn_1k)

## ?column?
## 1 0.231148

```

## Q. (2)

(a)

```
qn_4a2a <- c("select day from
(select vol, date_part('dow', retdate) as day
from stocks2016.d2010)q1
group by day
order by sum(vol) desc
limit 1;")
dbGetQuery(first_database,qn_4a2a)
```

```
##    day
## 1    2
```

(b)

```
qn_4a2b <- c("select avg(vol), date_part('dow',retdate::date)
as daysofweek from stocks2016.d2010
group by daysofweek order by avg desc
limit 1;")
dbGetQuery(first_database,qn_4a2b)
```

```
##      avg daysofweek
## 1 1316851          4
```

(c)

```
qn_4a2c <- c("select avg(cast(ret as numeric)),date_part('month', retdate::date) as months,
date_part('dow', retdate::date) as daysofweek from stocks2016.d2010
where ret != 'B' and ret !='C'
group by daysofweek, months order by avg desc
limit 1;
")
dbGetQuery(first_database,qn_4a2c)
```

```
##      avg months daysofweek
## 1 0.0111189      5          1
```

(d)

```
qn_4a2d <- c("select
case
when date_part('dow', retdate) = 1 then 'Monday'
when date_part('dow', retdate) = 2 then 'Tuesday'
when date_part('dow', retdate) = 3 then 'Wednesday'
when date_part('dow', retdate) = 4 then 'Thursday'
when date_part('dow', retdate) = 5 then 'Friday'
end as day,
avg(case when vol between 1000000 and 2000000 then vol end) as C2,
avg(case when vol < 1000000 or vol > 2000000 then vol end) as C3
from stocks2016.d2010
group by day;")
dbGetQuery(first_database,qn_4a2d)
```

```
##      day      c2      c3
## 1 Thursday 1417846 1309585
## 2 Wednesday 1418952 1268041
```

```

## 3    Monday 1418921 1111516
## 4    Tuesday 1420261 1287924
## 5    Friday 1416594 1306717

```

## 4B: Aggregation and Dates

**Q. (1)**

(a)

```

qn_2a <- c("select quarter from
           (select case
                     when date_part('quarter', retdate) = 1 then 'Q1'
                     when date_part('quarter', retdate) = 2 then 'Q2'
                     when date_part('quarter', retdate) = 3 then 'Q3'
                     when date_part('quarter', retdate) = 4 then 'Q4'
                     else null
                   end as quarter
               from stocks2016.d2010) as InnerQ
           group by 1
           order by count(quarter) desc limit 1;")
dbGetQuery(first_database,qn_2a)

```

```

##     quarter
## 1      Q4

```

(b)

```

qn_2b <- ("select permno, max(abs(prc)) as max_price from stocks2016.d2010
           where prc is not null
           group by permno
           order by 2 desc;")
dbGetQuery(first_database,qn_2b)

```

```

##      permno    max_price
## 1    17778 125612.00000
## 2    83443  3476.00000
## 3    36281  2006.00000
## 4    79785   753.00000
## 5    90857   714.16998
## [ reached getOption("max.print") -- omitted 7036 rows ]

```

(c)

```

qn_2c <- ("select permno, case
             when max_price > 100 then 1
             when max_price between 50 and 100 then 2
             when max_price < 50 then 3
             end as DFlag
           from
             (select permno, max(abs(prc)) as max_price
              from stocks2016.d2010
              where prc is not null
              group by permno) as InnerQ;")
dbGetQuery(first_database,qn_2c)

```

```

##      permno dflag
## 1     83799    1
## 2     85796    3
## 3     88292    2
## 4     92223    3
## 5    12310    3
## [ reached getOption("max.print") -- omitted 7036 rows ]

```

(d)

```

qn_2d <- ("select case
  when max_price > 100 then 1
  when max_price between 50 and 100 then 2
  when max_price < 50 then 3
  end as DFlag, count(1) as num_of_permnos
from
  (select permno, max(abs(prc)) as max_price
  from stocks2016.d2010
  where prc is not null
  group by permno) as InnerQ
group by 1;")
dbGetQuery(first_database,qn_2d)

```

```

##      dflag num_of_permnos
## 1        1           166
## 2        2           918
## 3        3          5957

```

(e)

```

qn_2e <- c("select count(case
  when max_price > 100 then 1
  else null
  end) as count_dflag_1,
count(case
  when max_price between 50 and 100 then 1
  else null
  end) as count_dflag_2,
count(case
  when max_price < 50 then 1
  else null
  end) as count_dflag_3
from
  (select permno, max(abs(prc)) as max_price
  from stocks2016.d2010
  where prc is not null
  group by permno) as InnerQ;")
dbGetQuery(first_database,qn_2e)

```

```

##      count_dflag_1 count_dflag_2 count_dflag_3
## 1            166           918          5957

```

(f)

```

qn_2f <- c("select date_part('month', retdate) as month, count(distinct retdate)
from stocks2016.d2010
group by 1;")

```

```

dbGetQuery(first_database, qn_2f)

##      month count
## 1      1    19
## 2      2    19
## 3      3    23
## 4      4    21
## 5      5    20
## 6      6    22
## 7      7    21
## 8      8    22
## 9      9    21
## 10     10   21
## 11     11   21
## 12     12   22

(g)

qn_2g <- c("select permno, max(abs(prc)) - min(abs(prc)) as price_diff
from stocks2016.d2010
where prc is not null and date_part('month', retdate) = 12
group by permno
having count(distinct retdate) = 22;")
dbGetQuery(first_database, qn_2g)

##      permno price_diff
## 1    10001    0.45000
## 2    10002    0.37000
## 3    10025    1.97000
## 4    10026    2.78000
## 5    10028    0.85000
##  [ reached getOption("max.print") -- omitted 6540 rows ]

(h)

qn_2h <- c(
  select avg(max_price - min_price) as avg_price_diff
  from
    (select permno, max(abs(prc)) as max_price, min(abs(prc)) as min_price
     from stocks2016.d2010
     where prc is not null and date_part('dow', retdate) = 2
       and date_part('month', retdate) = 1
      group by permno
     having count(distinct retdate) = 4) as InnerQ;
)
dbGetQuery(first_database, qn_2h)

##  avg_price_diff
## 1    2.580318

(i)

qn_2i <- c("select avg(max_price - min_price) as avg_price_diff
from
  (select permno, max(abs(prc)) as max_price, min(abs(prc)) as min_price
   from stocks2016.d2010
   where prc is not null
     and date_part('dow', retdate) = 3 and date_part('month', retdate) = 1
  ) as InnerQ;
")

```

```

        group by permno
        having count(distinct retdate) = 4) as InnerQ;")
dbGetQuery(first_database,qn_2i)

##    avg_price_diff
## 1      3.384888

(j)
qn_2j <- c("select avg(abs(prc)) as avg_price
from stocks2016.d2010
where permno in
(select permno
from stocks2016.d2010
where prc is not null and date_part('dow', retdate) = 2
and date_part('month', retdate) = 1
group by permno
having count(distinct retdate) = 4)
and date_part('dow', retdate) = 2 and date_part('month', retdate) = 1;")
dbGetQuery(first_database,qn_2j)

##    avg_price
## 1    37.4144

(k)
qn_2k <- c("select
count(case when num_tues = 4 and num_wed = 4
and avg_price_tues > avg_price_wed then 1 end) as days_tues_greater,
count(case when num_tues = 4 and num_wed = 4
and avg_price_tues < avg_price_wed then 1 end) as days_tues_lower
from
(select permno,
count(case when date_part('month', retdate) = 1
and date_part('dow', retdate) = 2 then 1 end) as num_tues,
count(case when date_part('month', retdate) = 1
and date_part('dow', retdate) = 3 then 1 end) as num_wed,
avg(case when date_part('dow', retdate) = 2
and date_part('month', retdate) = 1 then prc end) as avg_price_tues,
avg(case when date_part('dow', retdate) = 3
and date_part('month', retdate) = 1 then prc end) as avg_price_wed
from stocks2016.d2010
where prc is not null
group by permno) as InnerQ;")
dbGetQuery(first_database,qn_2k)

##    days_tues_greater days_tues_lower
## 1                  3322                 3182

```

## Q. (2)

(a)

```

qn_4b2a <- c("select avg(speakhr) from cls.traffic
where county ilike 'a%'"

```

```

and swpeakhr is not null;")
dbGetQuery(first_database,qn_4b2a)

##      avg
## 1 7238.977

(b)
qn_4b2b <- c("select county, avg(swpeakhr) as avgpeak
from cls.traffic
where county like 'A%'
group by county;")
dbGetQuery(first_database,qn_4b2b)

##      county    avgpeak
## 1  Amador   865.6098
## 2  Alpine    294.6667
## 3 Alameda  8996.0577

(c)
qn_4b2c <- c("select avg(swpeakhr) from cls.traffic
where county ilike 'a%' or county ilike 'm%';")
dbGetQuery(first_database,qn_4b2c)

##      avg
## 1 4098.753

(d)
qn_4b2d <- c("select avg_A, avg_M from
(select
avg(case when county ilike 'a%' then swpeakhr end) as avg_A,
avg(case when county ilike 'm%' then swpeakhr end) as avg_M
from cls.traffic)q1;")
dbGetQuery(first_database,qn_4b2d)

##      avg_a    avg_m
## 1 7238.977 2378.797

(e) Skip

(f)
qn_4b2f <- c("select count(distinct routeno) from cls.traffic
where swpeakhr > 5000 or nepeakhr < 2000
group by county;")
dbGetQuery(first_database,qn_4b2f)

##      count
## 1      16
## 2       4
## 3       6
## 4       6
## 5       4
## 6       3
## 7       8
## 8       4
## 9       5

```

```

## 10      13
## [ reached getOption("max.print") -- omitted 48 rows ]
(g)
qn_4b2g <- c("select county, routeno, avg(nepeakhr/swpeakhr) from cls.traffic
group by county, routeno;")
dbGetQuery(first_database,qn_4b2g)

##           county routeno      avg
## 1        Placer     89 0.5714286
## 2     Santa Clara     87 0.5000000
## 3         Kern    184 0.7000000
## [ reached getOption("max.print") -- omitted 516 rows ]

(h)
qn_4b2h <- c("select county from
(select county, routeno,
avg(nepeakhr/swpeakhr::float) as average
from cls.traffic
group by county, routeno)q1
where average is not null
group by county
order by sum(case when average < 1 then 1 else 0 end)*1.0/count(*) desc
limit 1;")
dbGetQuery(first_database,qn_4b2h)

##   county
## 1 Lassen

(i)
qn_4b2i <- c("select
count(case when swavgday_1 < 5000 then 1 else null end) as small_part,
count(case when swavgday_1 > 5000 then 1 else null end) as big_part
from(select county, routeno, avg(swavgday) as swavgday_1,
      avg(nepeakhr/swpeakhr) as ratio, count(routeno) as total_1
       from cls.traffic
      group by county, routeno)as q1
where ratio < 1;")
dbGetQuery(first_database,qn_4b2i)

##   small_part big_part
## 1          89       323

(j)
qn_4b2j <- c("select county, routeno, max(swpeakhr),
case
when max(swpeakhr) > 10000 then '>10,000'
when max(swpeakhr) > 5000 then '>5,000'
when max(swpeakhr) > 0 then '>0'
end as pkhr_type
from cls.traffic where
county in (select county from cls.traffic
group by county having max(swpeakhr)>15000)
and swpeakhr is not null")

```

```

group by county, routeno;")
dbGetQuery(first_database, qn_4b2j)

##           county routeno   max_pkhr_type
## 1      Santa Clara     87 13700    >10,000
## 2             Kern     184  2000        >0
## [ reached getOption("max.print") -- omitted 225 rows ]

```

## 4C: Information Schema and Exploring the Relationship between Returns and Dollar Volume

Q. (1)

(a)

```

qn_4c1a <- c("select count(1), data_type  from information_schema.columns
group by data_type;")
dbGetQuery(first_database, qn_4c1a)

```

```

##       count          data_type
## 1       8            anyarray
## 2       2              inet
## 3      31            "char"
## 4    33 timestamp with time zone
## 5      11            xid
## [ reached getOption("max.print") -- omitted 18 rows ]

```

(b)

```

qn_4c1b <- c("select  distinct(data_type)  from information_schema.columns;")
dbGetQuery(first_database, qn_4c1b)

```

```

##          data_type
## 1      anyarray
## 2          inet
## 3          "char"
## 4  timestamp with time zone
## 5            xid
## 6          name
## 7          bytea
## 8            date
## 9      double precision
## 10      pg_node_tree
## [ reached getOption("max.print") -- omitted 13 rows ]

```

(c)

```

qn_4c1c <- c("select count(1),data_type,table_schema from information_schema.columns
group by data_type,table_schema;")
dbGetQuery(first_database, qn_4c1c)

```

	count	data_type	table_schema
## 1	61	ARRAY	pg_catalog
## 2	198	oid	pg_catalog
## 3	11	pg_lsn	pg_catalog

```

## [ reached getOption("max.print") -- omitted 32 rows ]
(d)
qn_4c1d <- c("select table_schema
, count(case when data_type = '' then 1 end) as field_blank
, count(case when data_type ='anyarray' then 1 end) as field_array
, count(case when data_type ='inet' then 1 end) as field_inet
, count(case when data_type ='''char''' then 1 end) as field_char
, count(case when data_type ='timestamp with time zone' then 1 end) as field_time
, count(case when data_type ='xid' then 1 end) as field_xid
, count(case when data_type ='name' then 1 end) as field_name
, count(case when data_type ='bytea' then 1 end) as field_byte
, count(case when data_type ='date' then 1 end) as field_date
, count(case when data_type ='double precision' then 1 end) as field_doub
, count(case when data_type ='pg_node_tree' then 1 end) as field_pg_n
, count(case when data_type ='real' then 1 end) as field_real
, count(case when data_type ='interval' then 1 end) as field_inte
, count(case when data_type ='character varying' then 1 end) as field_char
, count(case when data_type ='abstime' then 1 end) as field_abst
, count(case when data_type ='bigint' then 1 end) as field_bigi
, count(case when data_type ='smallint' then 1 end) as field_smal
, count(case when data_type ='boolean' then 1 end) as field_bool
, count(case when data_type ='integer' then 1 end) as field_inte
, count(case when data_type ='ARRAY' then 1 end) as field_ARRA
, count(case when data_type ='oid' then 1 end) as field_oid
, count(case when data_type ='pg_lsn' then 1 end) as field_pg_l
, count(case when data_type ='regproc' then 1 end) as field_regp
, count(case when data_type ='text' then 1 end) as field_text
from information_schema.columns
group by 1;")

dbGetQuery(first_database,qn_4c1d)

```

	table_schema	field_blank	field_array	field_inet	field_char	
## 1	information_schema	0	0	0	0	
## 2	pg_catalog	0	8	2	0	
## 3	stocks2016	0	0	0	0	
## 4	cls	0	0	0	0	
	field_time	field_xid	field_name	field_byte	field_date	field_doub
## 1	3	0	3	0	0	0
## 2	30	11	124	2	0	8
## 3	0	0	0	0	4	19
## 4	0	0	0	0	0	2
	field_pg_n	field_real	field_inte	field_char	field_abst	field_bigi
## 1	0	0	0	579	0	0
## 2	11	9	2	0	2	150
## 3	0	0	0	18	0	10
## 4	0	0	0	9	0	0
	field_smal	field_bool	field_inte	field_arra	field_oid	field_pg_l
## 1	0	0	78	5	5	0
## 2	22	104	50	61	198	11
## 3	0	0	3	0	0	0
## 4	0	0	10	0	0	0
	field_regp	field_text				
## 1	0	0				

```

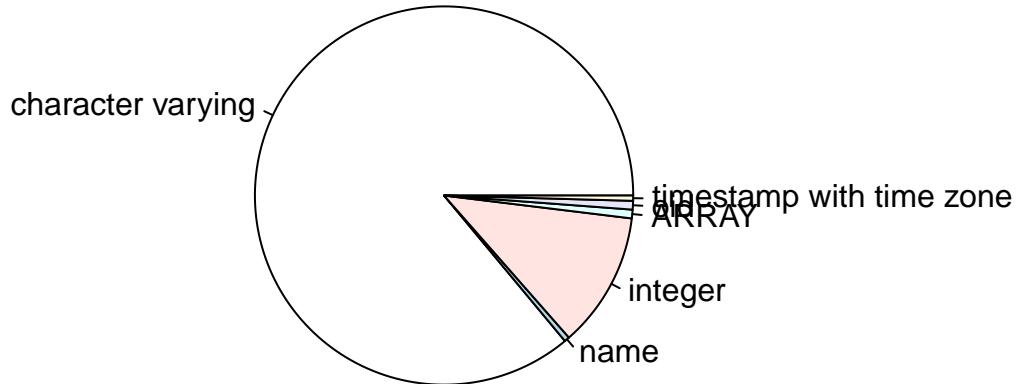
## 2      34      95
## 3      0       0
## 4      0       0

(e)

qn_4c1e <- c("select count(1),data_type,table_schema from information_schema.columns
where table_schema = 'information_schema'
group by data_type,table_schema;")
Query <- dbGetQuery(first_database,qn_4c1e)
Query

##   count          data_type      table_schema
## 1    579    character varying information_schema
## 2     3           name  information_schema
## 3    78           integer information_schema
## [ reached getOption("max.print") -- omitted 3 rows ]
pie(unlist(Query[1]),unlist(Query[2]))

```



## Q. (2)

(a)

```

qn_4c2a <- c("SELECT round(cast(ret as numeric),5)*100 as return,
(abs(vol*prc::bigint)/1000)*1000 as dollar_volume  FROM stocks2016.d2010
where ret != 'C' and ret != 'B' and md5(permno::varchar(100)) like '0%';")
new_data <- dbGetQuery(first_database,qn_4c2a)
new_data

```

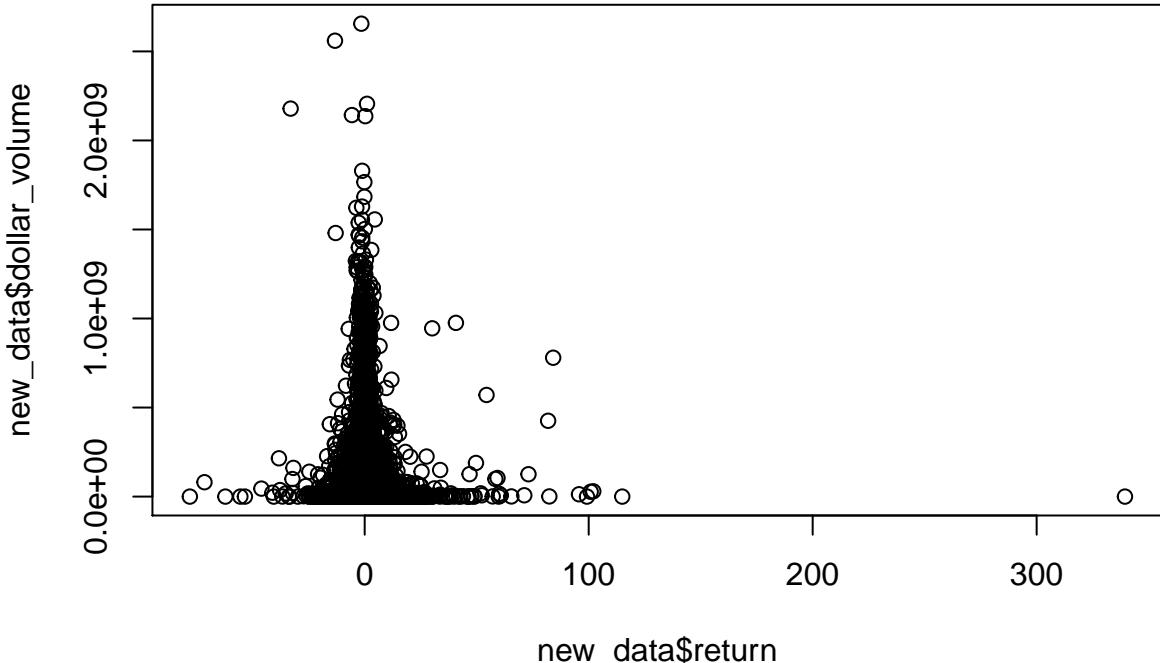
```

##          return dollar_volume
## 1      3.960      228000
## 2     -4.762      3719000
## 3      1.826      2507000
## 4     -3.252      2881000
## 5     -3.060      6228000
## [ reached getOption("max.print") -- omitted 113068 rows ]

```

(b)

```
plot(new_data$return,new_data$dollar_volume,type = 'p')
```



(c)

```
case <- lm(formula = new_data$dollar_volume ~ new_data$return)
print(summary(case))

##
## Call:
## lm(formula = new_data$dollar_volume ~ new_data$return)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30967023 -19746255 -18626641 -11813635 2635869311
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19934771    221839   89.861  <2e-16 ***
## new_data$return     32505     67192    0.484    0.629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 74560000 on 113071 degrees of freedom
## Multiple R-squared:  2.07e-06,   Adjusted R-squared:  -6.774e-06
## F-statistic: 0.234 on 1 and 113071 DF,  p-value: 0.6286
```

With  $R^2$  approximately 0, there is no apparent linear relationship between returns and volume traded.

(d)

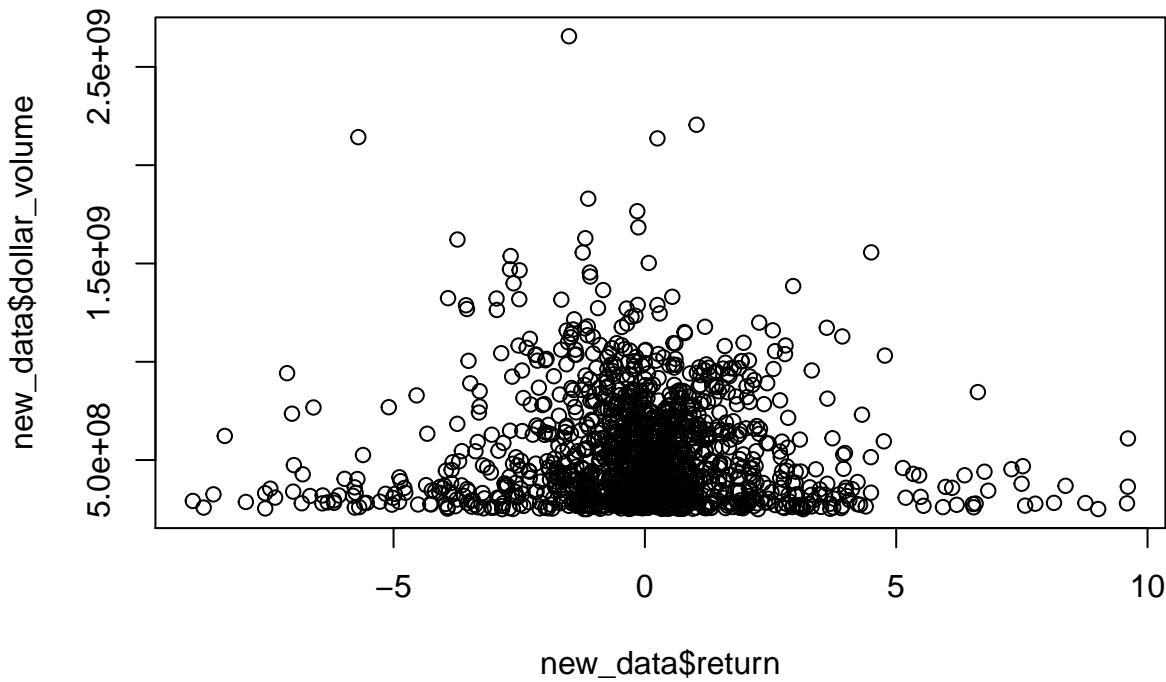
```
qn_4c2d <- c("SELECT round(cast(ret as numeric),5)*100 as return,
(abs(vol*prc::bigint)/1000)*1000 as dollar_volume FROM stocks2016.d2010
where (ret::numeric) between -.1 and .1
and (abs(vol*prc::bigint)/1000)*1000 > 250000000
and ret != 'C' and ret != 'B' and md5(permno::varchar(100)) like '0%';")
```

```

new_data <- dbGetQuery(first_database, qn_4c2d)
new_data

##      return dollar_volume
## 1    9.608   364568000
## 2    9.591   279806000
## 3    9.019   250845000
## 4   -6.594   767464000
## 5    6.756   440681000
## [ reached getOption("max.print") -- omitted 1451 rows ]
plot(new_data$return, new_data$dollar_volume, type = 'p')

```



There does not seem to be a linear relationship between returns and dollar volume.

(e)

```

case <- lm(formula = new_data$dollar_volume ~ new_data$return)
print(summary(case))

##
## Call:
## lm(formula = new_data$dollar_volume ~ new_data$return)
##
## Residuals:
##      Min       1Q     Median       3Q      Max 
## -320077944 -213342344  -79104796  129095177 2115970711 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 531781029    7258666  73.262  <2e-16 ***
## new_data$return -5300173    3422424  -1.549    0.122  
## ---    
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

## 
## Residual standard error: 276900000 on 1454 degrees of freedom
## Multiple R-squared:  0.001647,   Adjusted R-squared:  0.0009601
## F-statistic: 2.398 on 1 and 1454 DF,  p-value: 0.1217

```

With  $R^2$  approximately 0 there is no apparent linear relationship between returns and dollar volume.

(f)

```

qn_4c2f <- c("select avg(return^2)-avg(return)^2 as variance_return,
avg(dollar_volume^2)-avg(dollar_volume)^2 as dollar_value_variance
from (SELECT round(cast(ret as numeric),5)*100 as return,
(abs(vol*prc::bigint)/1000)*1000 as dollar_volume  FROM stocks2016.d2010
where ret != 'C' and ret != 'B'
and md5(permno::varchar(100)) like '0%') as innerq;")
dbGetQuery(first_database,qn_4c2f)

```

```

##  variance_return dollar_value_variance
## 1      10.89011      5.559268e+15

```

(g)

```

qn_4c2g <- c("select avg(return * dollar_volume) - (avg(return)*avg(dollar_volume))
as covariance_of_return_dollar_value from
(SELECT round(cast(ret as numeric),5)*100 as return,
(abs(vol*prc::bigint)/1000)*1000 as dollar_volume  FROM stocks2016.d2010
where ret != 'C'
and ret != 'B' and md5(permno::varchar(100)) like '0%') as innerq;")
dbGetQuery(first_database,qn_4c2g)

```

```

##  covariance_of_return_dollar_value
## 1                  353978.4

```