

# Descriptive Analysis of Cohort

## Contents

Basic Statistics . . . . .	1
Time Spent Taking Only Antihypertensives of a Single Group (Windows) . . . . .	1
Kinds of Antihypertensive Drugs . . . . .	4
Temporal Analysis of Blood Pressure Measurements . . . . .	5
Diagnoses Codes Prior to First Prescription of a Single Antihypertensive, by Drug Group . . . . .	12
Number of Drug Orders . . . . .	14

## Basic Statistics

```
cohort = qfetch(db, "  
select * from new_cohort  
")  
  
dim(cohort)
```

```
## [1] 3264 4
```

```
summary(cohort)
```

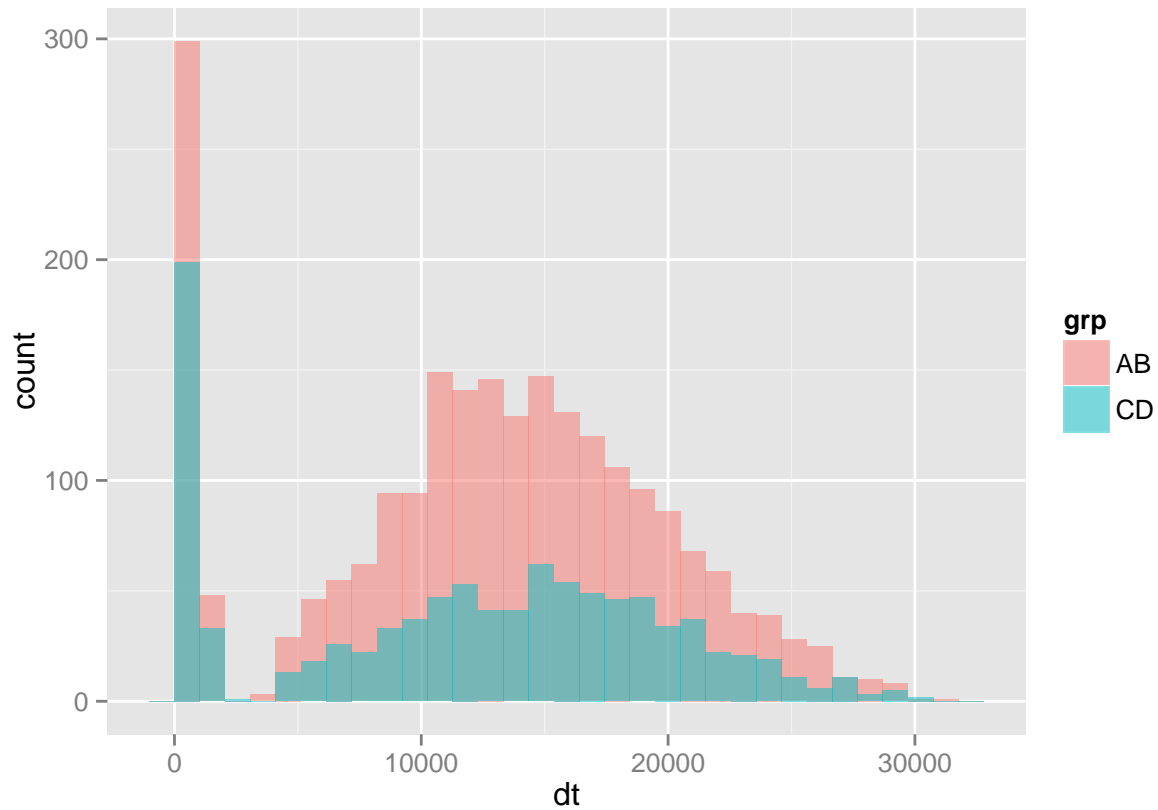
```
## patient_id grp t0 tf  
## Min. : 930 AB:2271 Min. : 5660 Min. : 8368  
## 1st Qu.: 503756 CD: 993 1st Qu.:17979 1st Qu.:36500  
## Median :1001090 Median :21906 Median :36500  
## Mean :1006592 Mean :21642 Mean :34013  
## 3rd Qu.:1506186 3rd Qu.:25588 3rd Qu.:36500  
## Max. :2010926 Max. :32461 Max. :36500
```

## Time Spent Taking Only Antihypertensives of a Single Group (Windows)

The upshot of this is that there are two clearly separate usage patterns: 1. those who were on one class of drugs for some time and then got another within a few months to a few years (15%). Are these people switchers, adders or did they take the first or second drugs for something unrelated to HT? 2. those who never took a drug of the other class after taking their first drug (85%) are these people well-controlled hypertensives, or did they take the first drug for something unrelated to HT?

Seems to be no significant age difference in the first drug start time between the two drug groups or two usage patterns or the combination.

```
cohort$dt = cohort$tf- cohort$t0  
  
ggplot(cohort, aes(x=dt, fill=grp)) +  
  geom_histogram(alpha=.5, position="identity")
```



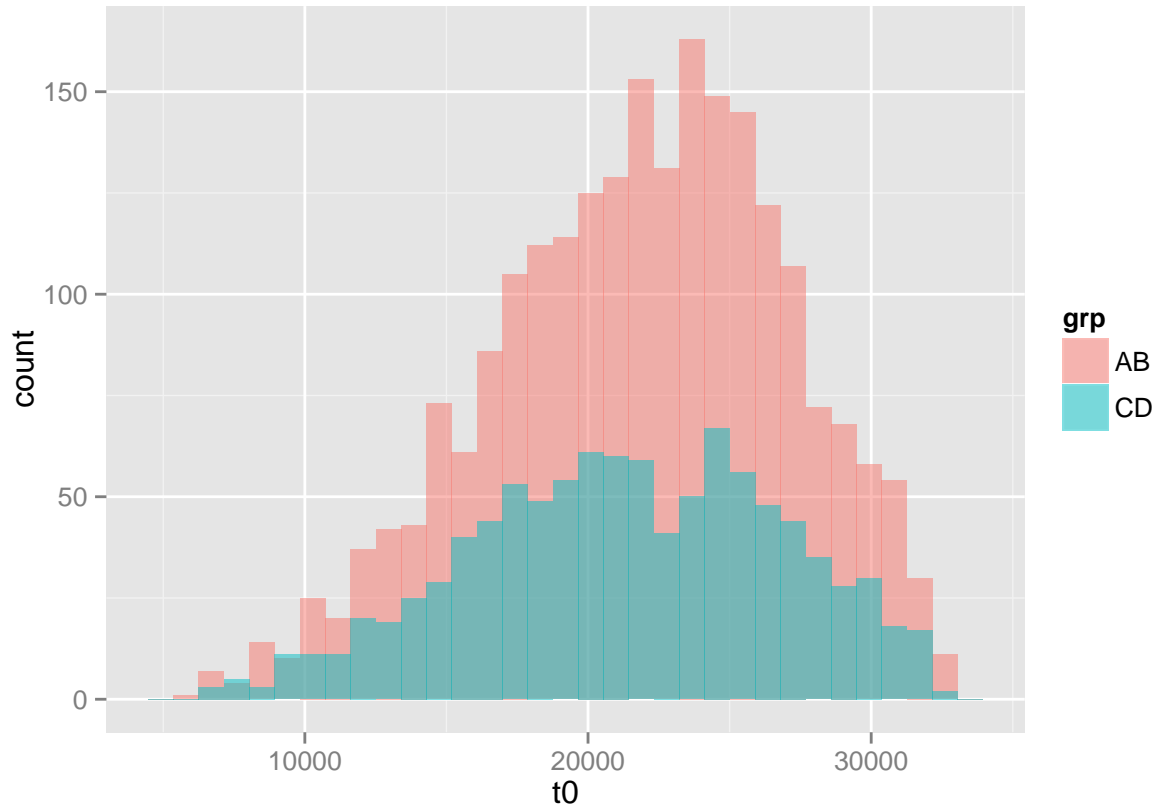
```
summary(cohort$tf[cohort$dt>2000])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  27500  36500   36500   36500  36500   36500
```

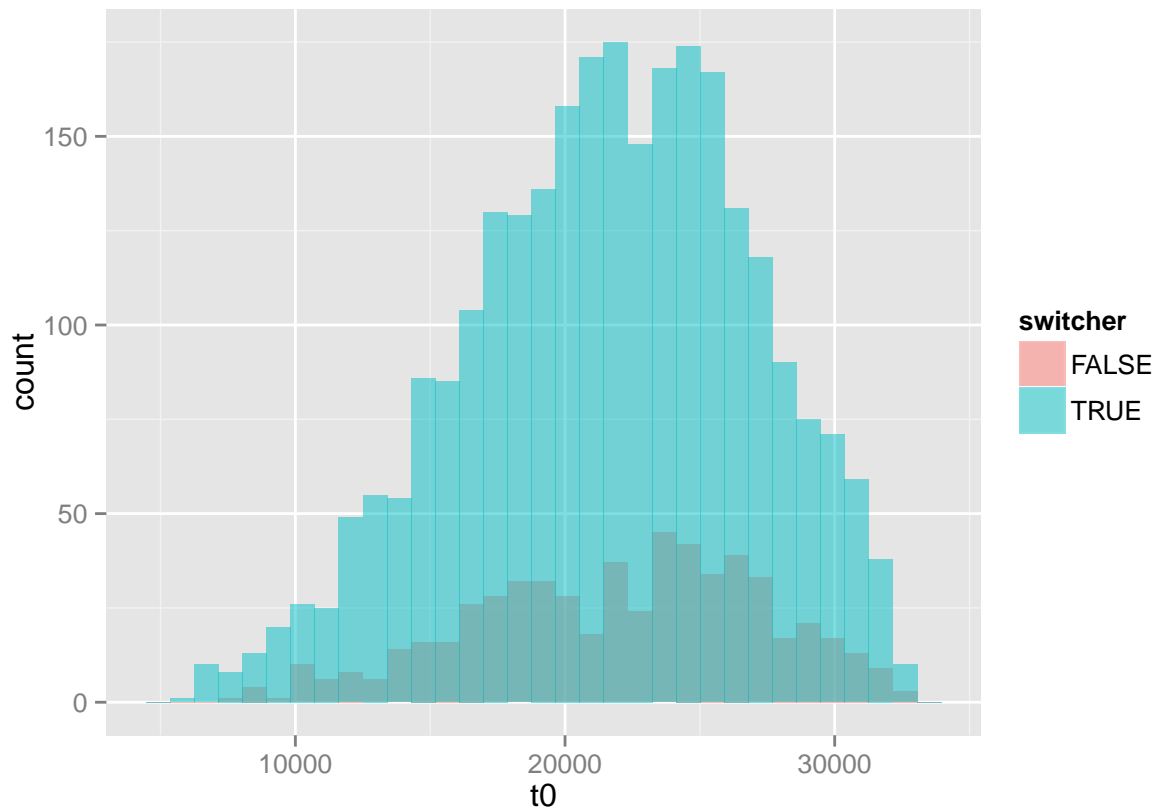
```
# Since dt stratifies the group pretty well, let's make it a factor
# and see what it does to the analysis
cohort$switcher=as.factor(cohort$tf==36500)
sum(cohort$switcher==TRUE)/length(cohort$patient_id)
```

```
## [1] 0.8223039
```

```
ggplot(cohort, aes(x=t0, fill=grp)) +
  geom_histogram(alpha=.5, position="identity")
```



```
ggplot(cohort, aes(x=t0, fill=switcher)) +  
  geom_histogram(alpha=.5, position="identity")
```



## Kinds of Antihypertensive Drugs

Number of different drugs at first antihypertensive perscription

```
kable(qfetch(db, "  
SELECT COUNT(patient_id) AS n_patients, n AS n_initial_drugs FROM (  
SELECT COUNT(DISTINCT o0.ingr_set_id) AS n, patient_id FROM (  
SELECT o.*, c.grp  
from new_cohort c  
JOIN drug_orders o  
  ON c.patient_id = o.patient_id  
  AND c.t0 = o.age_at_start_time_in_days  
  ) o0  
GROUP BY o0.patient_id) cnt  
GROUP BY cnt.n  
"))
```

<u>n_patients</u>	<u>n_initial_drugs</u>
3077	1
183	2
4	3

Multi-antihypertensive usage over lifetime

```
kable(qfetch(db, "  
SELECT n_drugs, COUNT(DISTINCT patient_id) AS n_patients  
FROM  
  (SELECT COUNT(DISTINCT o.ingr_set_id) AS n_drugs, o.patient_id  
  from new_cohort c  
  JOIN drug_orders o  
    ON c.patient_id = o.patient_id  
  GROUP BY o.patient_id  
  ) AS inter  
GROUP BY n_drugs  
ORDER BY n_drugs  
"))
```

<u>n_drugs</u>	<u>n_patients</u>
1	1905
2	896
3	293
4	107
5	44
6	10
7	7
8	2

## Multi-antihypertensive usage over first group window

During the window of having taken drugs only from one antihypertensive drug group, most patients only took one antihypertensive. These results don't vary when stratifying by drug group.

```
kable(qfetch(db, "
SELECT n_drugs, COUNT(DISTINCT patient_id) AS n_patients
FROM
  (SELECT COUNT(DISTINCT o.ingr_set_id) AS n_drugs, o.patient_id
   from new_cohort c
   JOIN drug_orders o
     ON c.patient_id = o.patient_id
   WHERE o.age_at_start_time_in_days < c.tf
   GROUP BY o.patient_id
  ) AS inter
GROUP BY n_drugs
ORDER BY n_drugs"))
```

n_drugs	n_patients
1	2309
2	808
3	129
4	15
5	3

## Temporal Analysis of Blood Pressure Measurements

Most measurements happen right before the first drug and about 180 days after the first drug, consistent with the behavior we would expect from physicians diagnosing high blood pressure and prescribing an antihypertensive.

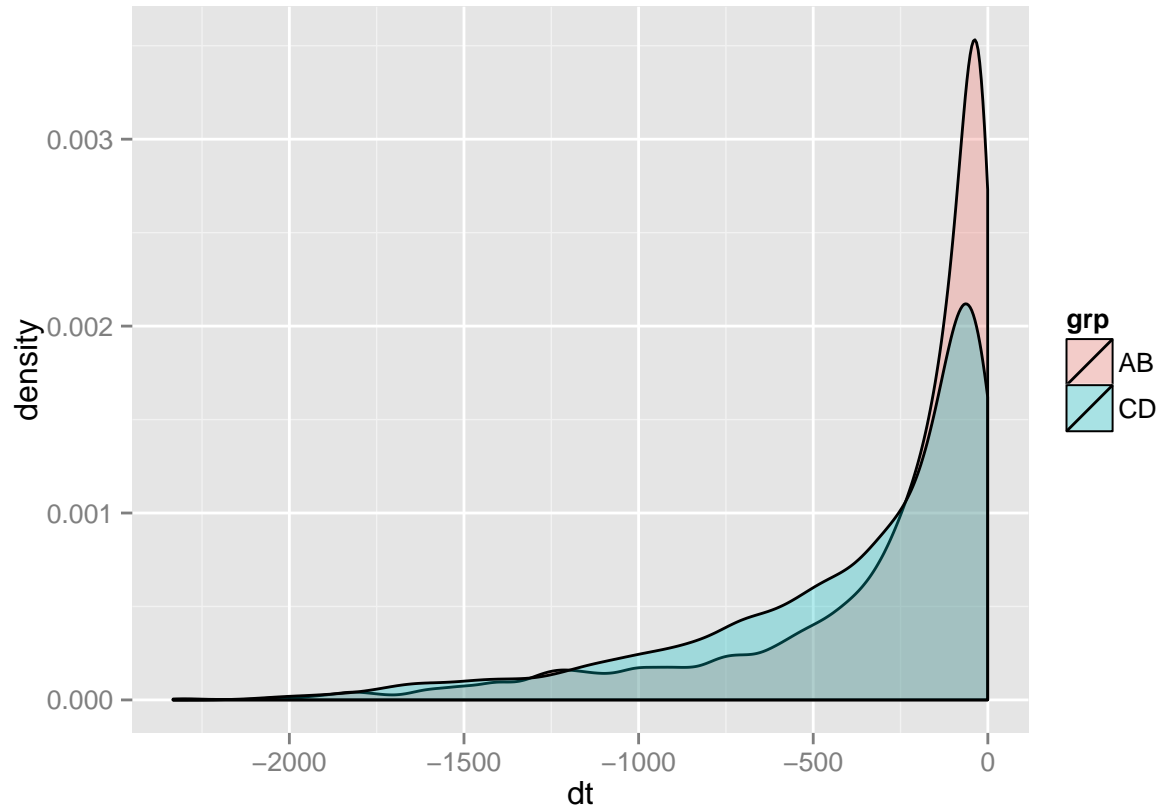
### Distribution in time BEFORE the first antihypertensive

Get all bp measurements and the latest bp measurement per patient prior to the first antihypertensive

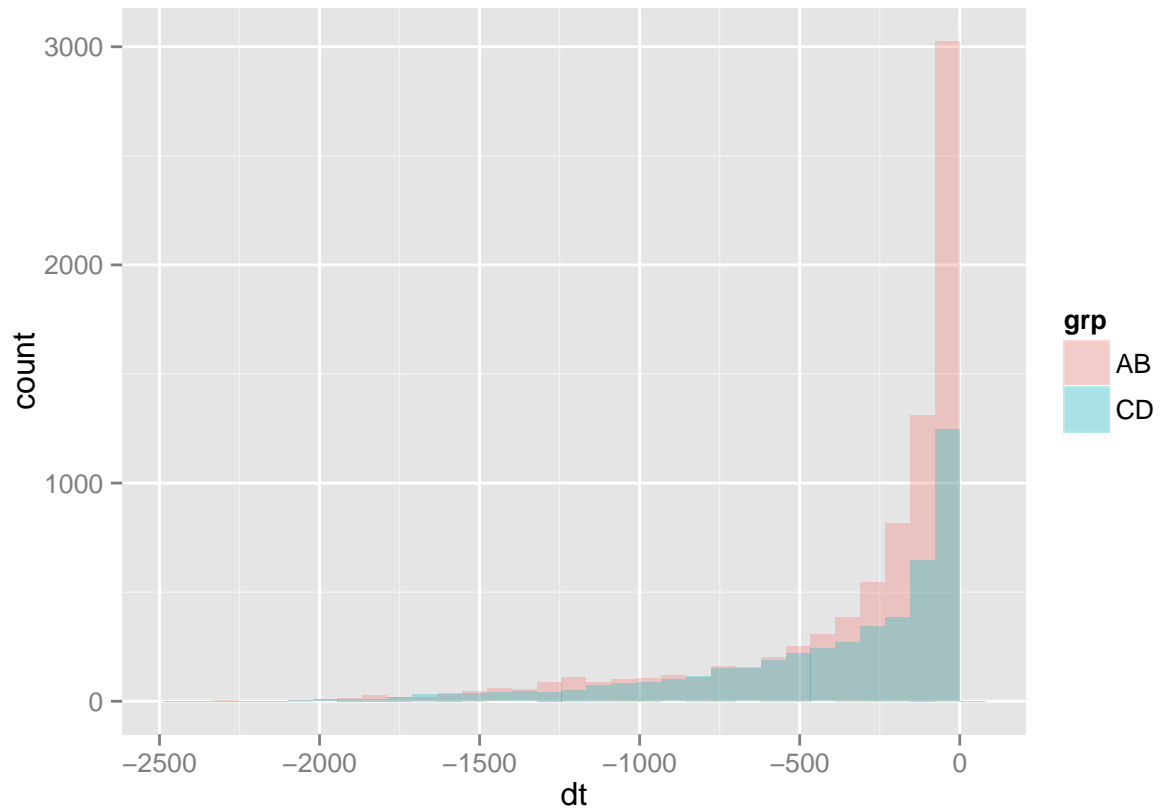
```
pre_bp = qfetch(db, "
select bpv.t-bpv.t0 as dt, c.patient_id, c.grp
from pre_drug_bp_visits bpv
join new_cohort c
on c.patient_id = bpv.patient_id
")
latest_pre_bp = qfetch(db, "
select max(bpv.t-bpv.t0) as dt, c.patient_id, c.grp
from pre_drug_bp_visits bpv
join new_cohort c
on c.patient_id = bpv.patient_id
group by c.patient_id
")
```

Times of all previous blood pressure measurements relative to the first perscription of a single antihypertensive for patients in our cohort

```
ggplot(pre_bp, aes(x=dt, fill=grp)) +  
  geom_density(position="identity", alpha=0.3)
```

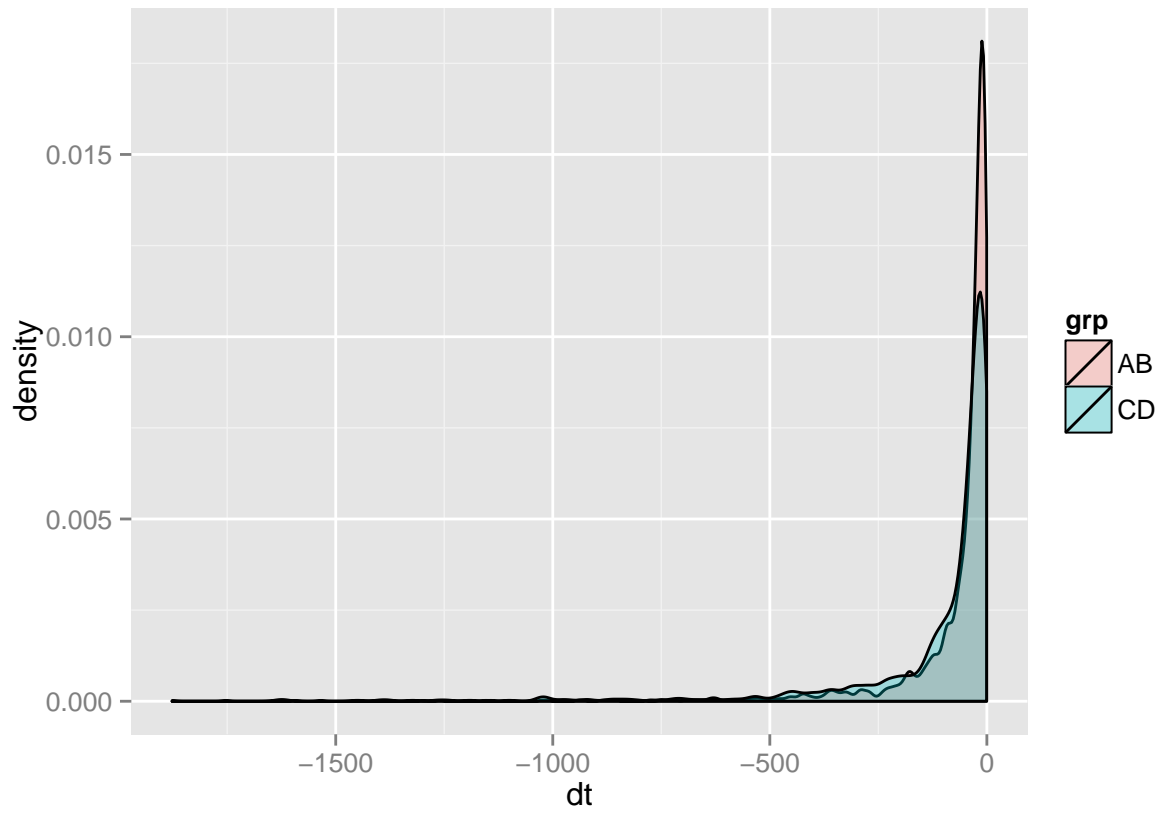


```
ggplot(pre_bp, aes(x=dt, fill=grp)) +  
  geom_histogram(position="identity", alpha=0.3)
```

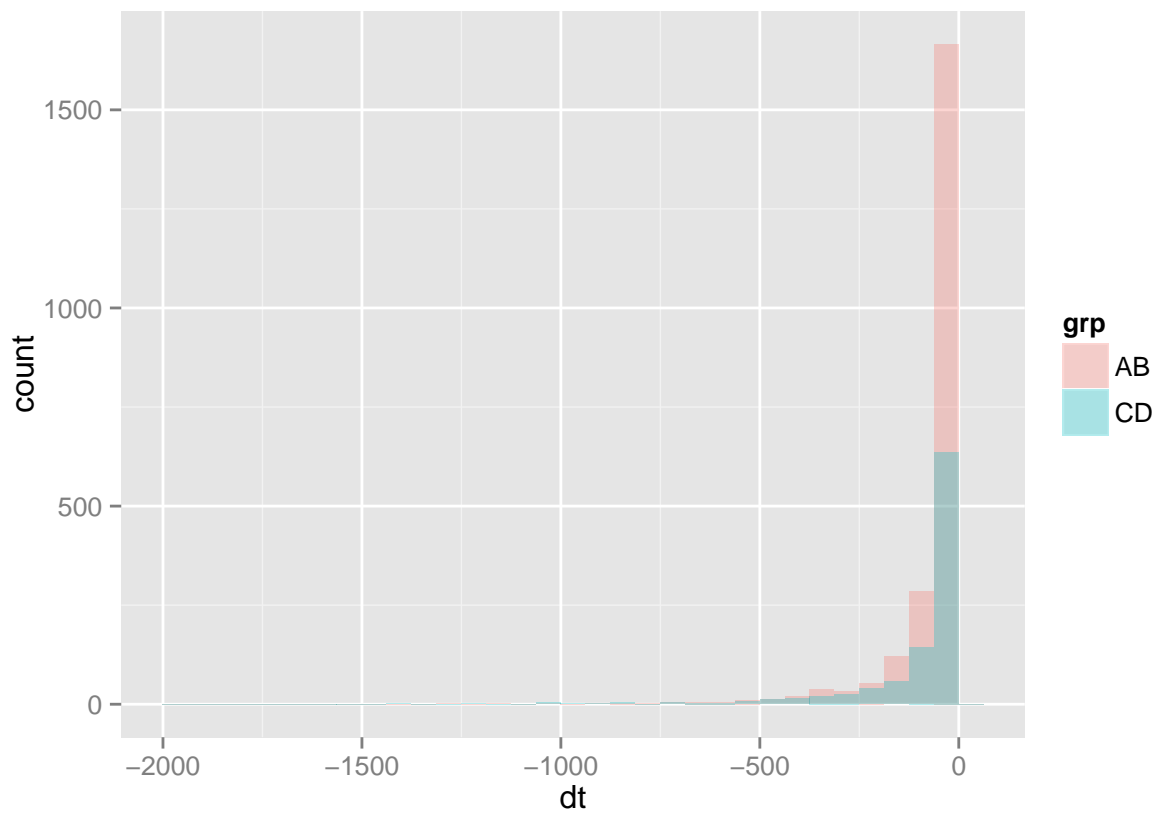


Times of latest blood pressure measurement prior to the first perscription of a single antihypertensive for each patient in our cohort (relative to the time of first perscription)

```
ggplot(latest_pre_bp, aes(x=dt, fill=grp)) +  
  geom_density(position="identity", alpha=0.3)
```



```
ggplot(latest_pre_bp, aes(x=dt, fill=grp)) +  
  geom_histogram(position="identity", alpha=0.3)
```





Not a big difference between the two groups. CD has a longerish tail, which is interesting, but probably not worth looking into too much.

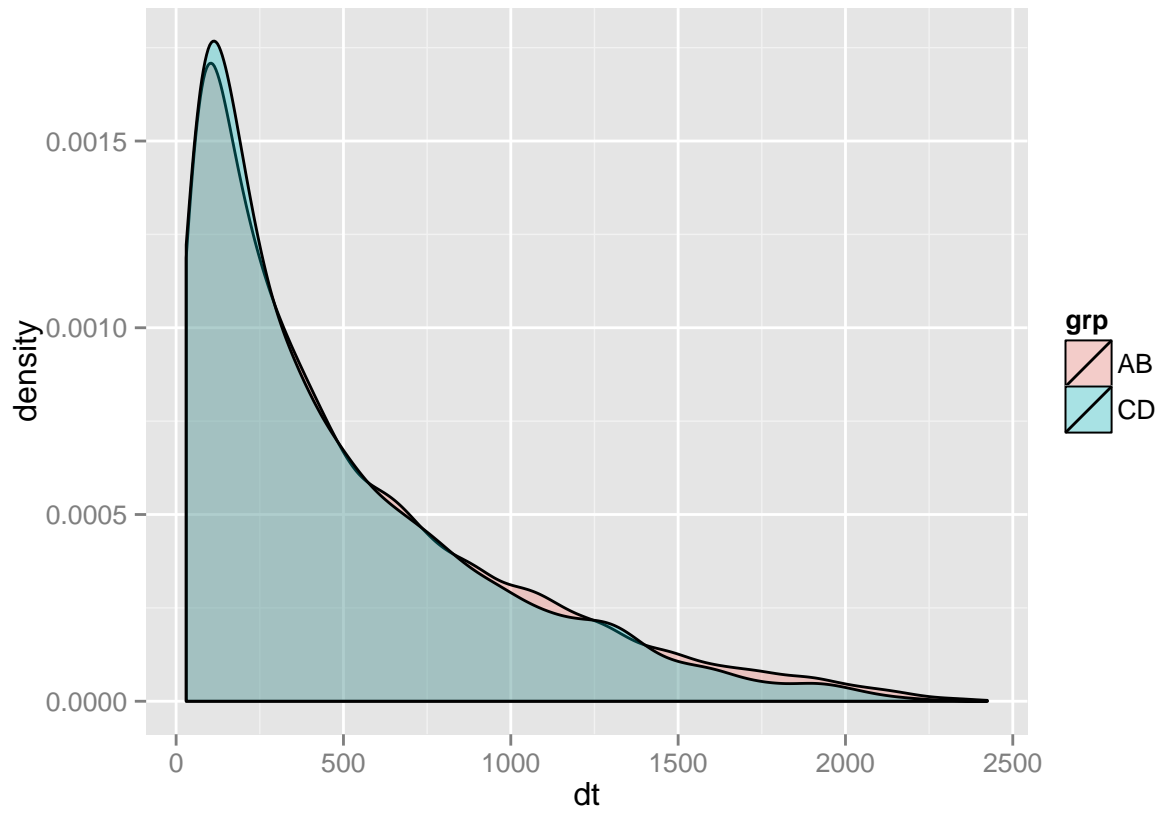
### Distribution in time AFTER the first antihypertensive

Get all bp measurements and the earliest bp measurement per patient after the first antihypertensive

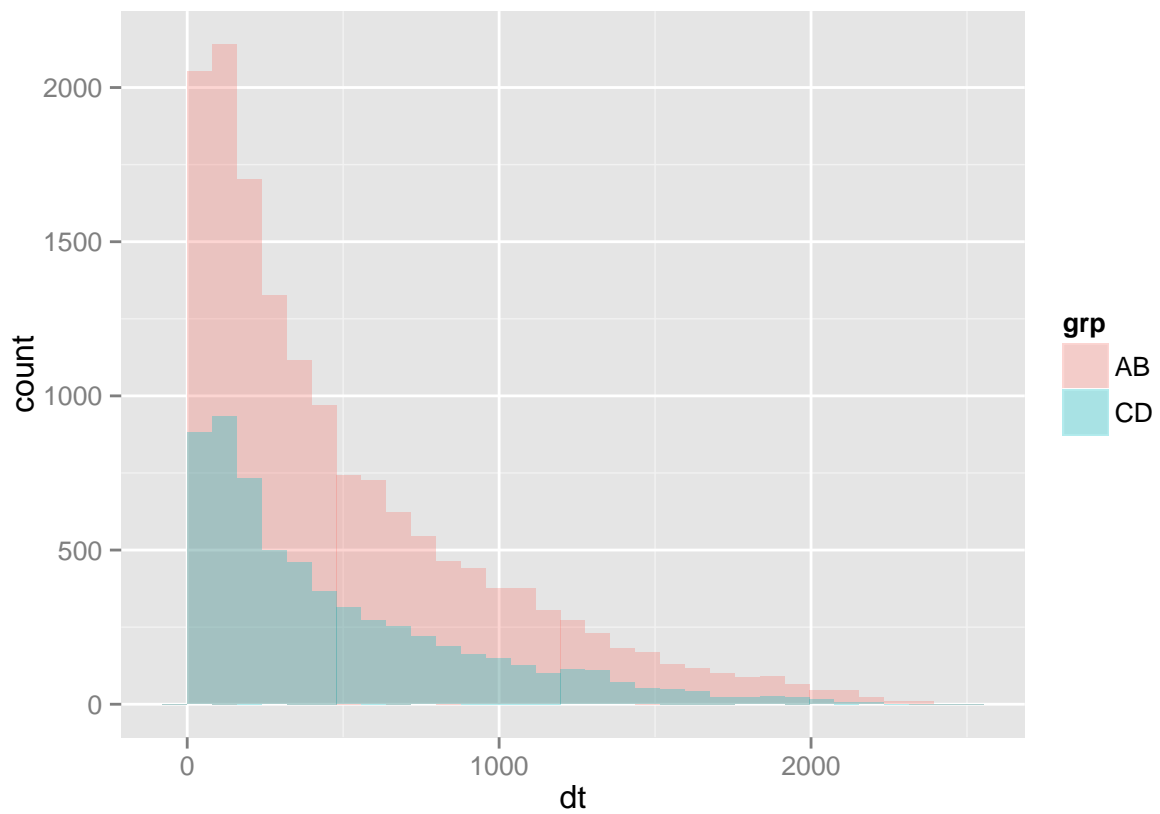
```
post_bp = qfetch(db, "  
select bpv.t-bpv.t0 as dt, c.patient_id, c.grp  
from post_drug_bp_visits bpv  
join new_cohort c  
on c.patient_id = bpv.patient_id  
")  
  
earliest_post_bp = qfetch(db, "  
select min(bpv.t-bpv.t0) as dt, c.patient_id, c.grp  
from post_drug_bp_visits bpv  
join new_cohort c  
on c.patient_id = bpv.patient_id  
group by c.patient_id  
")
```

Times of all subsequent blood pressure measurements relative to the first perscription of a single antihypertensive for patients in our cohort

```
ggplot(post_bp, aes(x=dt, fill=grp)) +  
  geom_density(position="identity", alpha=0.3)
```

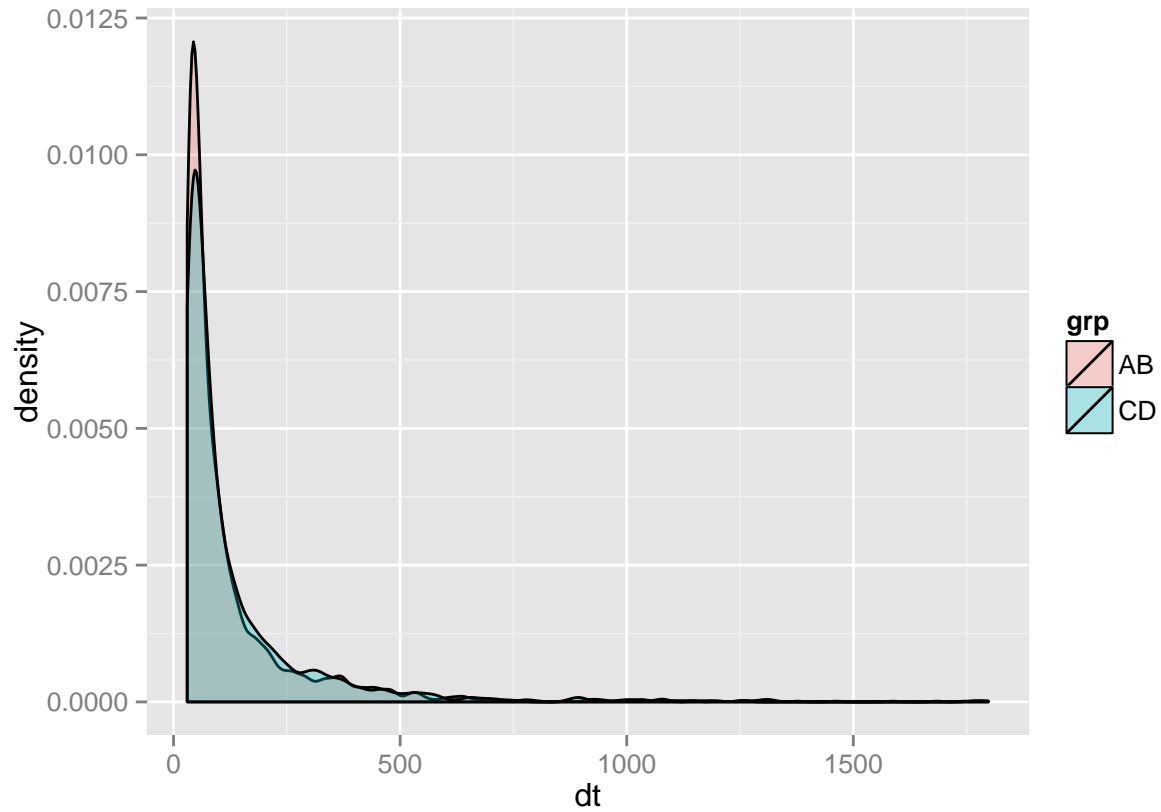


```
ggplot(post_bp, aes(x=dt, fill=grp)) +  
  geom_histogram(position="identity", alpha=0.3)
```

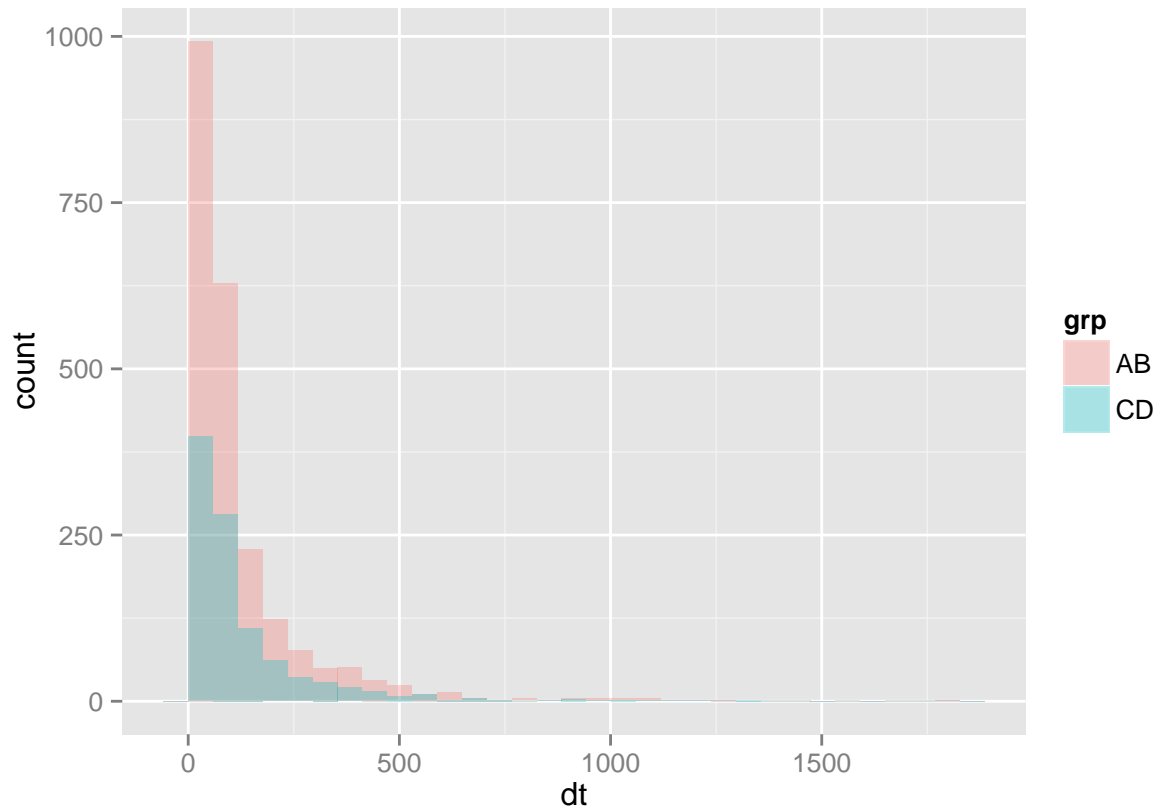


Times of earliest blood pressure measurement subsequent to the first perscription of a single antihypertensive for each patient in our cohort (relative to the time of first perscription)

```
ggplot(earliest_post_bp, aes(x=dt, fill=grp)) +  
  geom_density(position="identity", alpha=0.3)
```



```
ggplot(earliest_post_bp, aes(x=dt, fill=grp)) +  
  geom_histogram(position="identity", alpha=0.3)
```



## Diagnoses Codes Prior to First Prescription of a Single Antihypertensive, by Drug Group

### AB patients

```

dx = qfetch(db, "
SELECT COUNT(DISTINCT codes.patient_id) n, new_cohort.grp, codes.code
FROM codes
join new_cohort
ON codes.patient_id = new_cohort.patient_id
WHERE new_cohort.grp = 'AB'
AND codes.code IS NOT NULL
AND codes.code != ''
AND codes.c_type != 'CPT'
GROUP BY codes.code
ORDER BY n DESC
LIMIT 20
")
dx$str = sapply(dx$code, icd9Explain, doCondense=FALSE, brief=T)
kable(dx[,c('str', 'n')])

```

str	n
Hypertension NOS	2212
Hyperlipidemia NEC/NOS	934
DMII wo cmp nt st uncntr	564

str	n
Routine medical exam	542
Preop exam unspcf	359
Cor ath unsp vsl ntv/gft	327
Esophageal reflux	307
Chest pain NOS	296
Depressive disorder, not elsewhere classified	250
Obesity NOS	229
Lumbago	225
Joint pain-l/leg	215
Hypothyroidism NOS	209
Malaise and fatigue NEC	205
Obstructive sleep apnea	200
Cough	187
Anxiety state NOS	178
Pain in limb	176
Atrial fibrillation	169
Respiratory abnorm NEC	164

## CD patients

```

dx = qfetch(db, "
SELECT COUNT(DISTINCT codes.patient_id) n, new_cohort.grp, codes.code
FROM codes
join new_cohort
ON codes.patient_id = new_cohort.patient_id
WHERE new_cohort.grp = 'CD'
AND codes.code IS NOT NULL
AND codes.code != ''
AND codes.c_type != 'CPT'
GROUP BY codes.code
ORDER BY n DESC
LIMIT 20
")
dx$str = sapply(dx$code, icd9Explain, doCondense=FALSE, brief=T)
kable(dx[,c('str', 'n')])

```

str	n
Hypertension NOS	979
Routine medical exam	396
Hyperlipidemia NEC/NOS	298
Chest pain NOS	142
Screen mammogram NEC	139
Esophageal reflux	138
Cough	134
Lumbago	127
Pain in limb	125
Joint pain-l/leg	124
Preop exam unspcf	121
Malaise and fatigue NEC	120

str	n
Depressive disorder, not elsewhere classified	119
Elev bl pres w/o hypertn	113
Headache	112
Insomnia NOS	109
Backache NOS	107
Vaccin for influenza	106
Dermatitis NOS	106
Abdmnal pain unspcf site	106

## Number of Drug Orders

### Order Statuses of First Antihypertensive Orders

```
os = qfetch(db, "
SELECT COUNT(*) AS n, order_status FROM (
SELECT o.*, c.grp
FROM new_cohort c
JOIN drug_orders o
  ON c.patient_id = o.patient_id
  AND c.t0 = o.age_at_start_time_in_days
) o0
GROUP BY o0.order_status
ORDER BY n DESC
")
kable(os)
```

n	order_status
2459	Sent
1092	Discontinued
27	Completed
26	Suspend

Need to figure out what's going on with these "discontinued" orders, since I don't know if they can possibly be refills, considering they're the first ever order of an antihypertensive. Is it possible that they're a "formulary change" from a totally different drug, or an add-on to a prior drug?

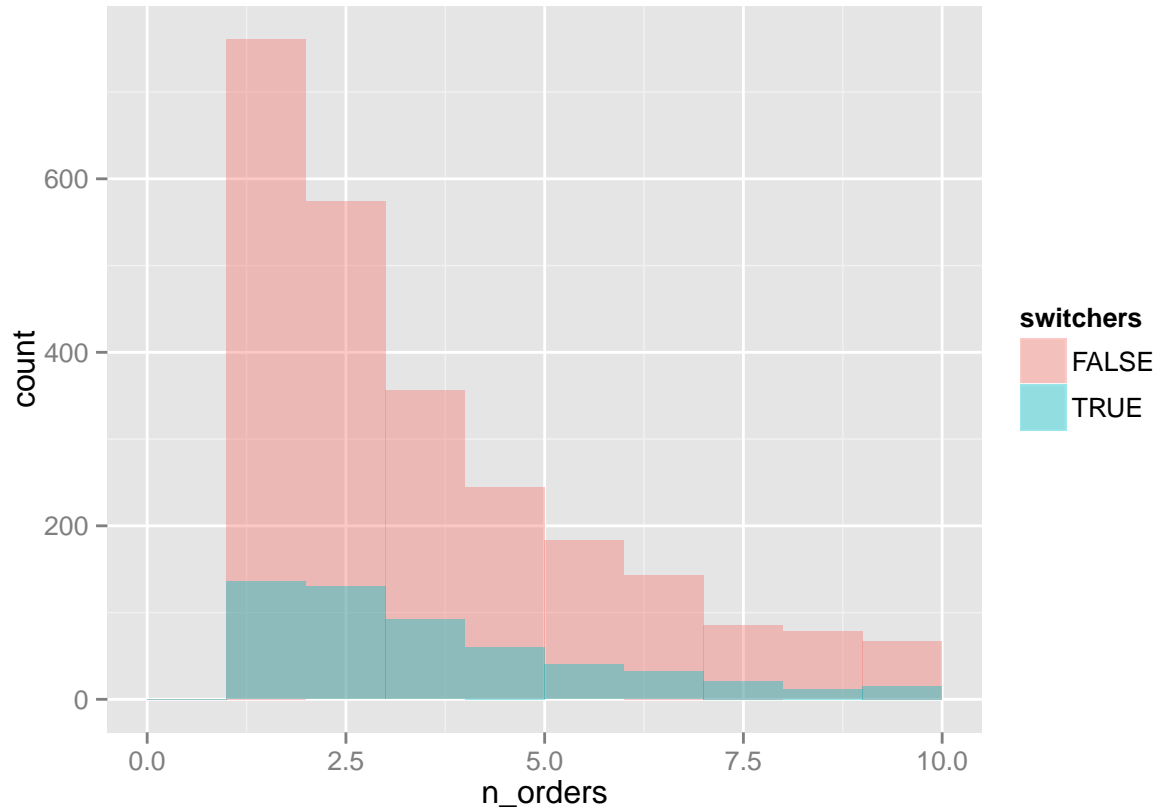
### Antihypertensive drug orders during the window

```
window_orders = qfetch(db, "
SELECT COUNT(*) AS n_orders, o.patient_id, c.grp, c.tf
FROM new_cohort c
```

```

JOIN drug_orders o
  ON c.patient_id = o.patient_id
WHERE o.age_at_start_time_in_days < c.tf
GROUP BY o.patient_id
")
window_orders$switchers = (window_orders$tf!=36500)
ggplot(window_orders, aes(x=n_orders, fill=switchers)) +
  geom_histogram(alpha=0.4, binwidth=1, position='identity') +
  xlim(0,10)

```



A large number of people only got one drug order during their window, and that doesn't correlate well to who switched and who stayed on the same drug group for life, according to our data. Perhaps the people who only got the single script started going elsewhere after their diagnosis? Or maybe the refills are infinite?