Exploration of data

Jon\_Minton

Monday, January 26, 2015

# Introduction

This document will show the results of some exploration of the data.

First, to load the R packages and datasets.

## Loading required package: ggplot2  
## Loading required package: lubridate  
## Loading required package: stringr  
## Loading required package: plyr  
##   
## Attaching package: 'plyr'  
##   
## The following object is masked from 'package:lubridate':  
##   
## here  
##   
## Loading required package: dplyr  
##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:plyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize  
##   
## The following objects are masked from 'package:lubridate':  
##   
## intersect, setdiff, union  
##   
## The following object is masked from 'package:stats':  
##   
## filter  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union  
##   
## Loading required package: tidyr  
## Loading required package: car

Now to look at data\_main

# Start with data\_main  
  
summary(data\_main)

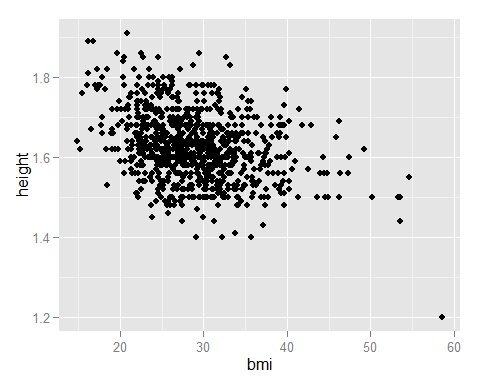
## X record\_number hosp\_number health\_centre   
## Min. : 1.0 Min. : 1.0 : 12 Min. :1.000   
## 1st Qu.:242.2 1st Qu.: 246.8 1005/08 : 2 1st Qu.:1.000   
## Median :483.5 Median : 570.5 1006/08 : 2 Median :1.000   
## Mean :483.5 Mean : 542.8 1083/08 : 2 Mean :1.276   
## 3rd Qu.:724.8 3rd Qu.: 814.2 1084/09 : 2 3rd Qu.:2.000   
## Max. :966.0 Max. :1064.0 11027/08: 2 Max. :2.000   
## NA's :2 (Other) :944 NA's :5   
## hospital\_visited num\_hospital\_visits health\_status age   
## Min. :1.000 Min. : 1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.: 4.000 1st Qu.:1.000 1st Qu.:4.000   
## Median :2.000 Median : 6.000 Median :2.000 Median :4.000   
## Mean :1.944 Mean : 6.309 Mean :1.768 Mean :4.427   
## 3rd Qu.:2.000 3rd Qu.: 8.000 3rd Qu.:2.000 3rd Qu.:5.000   
## Max. :2.000 Max. :16.000 Max. :4.000 Max. :8.000   
## NA's :8 NA's :35 NA's :25 NA's :24   
## health\_insurance\_status education marital\_status occupation   
## Min. :1.000 Min. :1.000 Min. :1.00 BUSINESSMAN: 92   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:2.00 DRIVER : 92   
## Median :1.000 Median :3.000 Median :2.00 TRADER : 65   
## Mean :1.009 Mean :3.124 Mean :1.88 : 47   
## 3rd Qu.:1.000 3rd Qu.:4.000 3rd Qu.:2.00 MASON : 35   
## Max. :2.000 Max. :5.000 Max. :5.00 MECHANIC : 24   
## NA's :10 NA's :14 NA's :26 (Other) :611   
## parity household\_size works occupation\_main  
## Min. :0.000 Min. : 1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.: 3.000 1st Qu.:1.000 1st Qu.:3.000   
## Median :2.000 Median : 4.000 Median :2.000 Median :3.000   
## Mean :2.042 Mean : 4.179 Mean :1.672 Mean :2.701   
## 3rd Qu.:3.000 3rd Qu.: 5.000 3rd Qu.:2.000 3rd Qu.:3.000   
## Max. :8.000 Max. :10.000 Max. :2.000 Max. :3.000   
## NA's :4 NA's :18 NA's :21 NA's :839   
## occupation\_main\_specific monthly\_income religion   
## :321 Min. :1.0 Min. :1.000   
## TRADER :215 1st Qu.:1.0 1st Qu.:2.000   
## SEAMSTRESS : 77 Median :1.0 Median :2.000   
## HAIRDRESSER: 48 Mean :1.6 Mean :1.902   
## trader : 40 3rd Qu.:2.0 3rd Qu.:2.000   
## CATERER : 19 Max. :4.0 Max. :4.000   
## (Other) :246 NA's :314 NA's :19   
## belief\_cannot\_control\_health belief\_things\_do\_affect\_pregnancy  
## Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:1.000   
## Median :2.000 Median :2.000   
## Mean :2.042 Mean :2.494   
## 3rd Qu.:3.000 3rd Qu.:4.000   
## Max. :5.000 Max. :5.000   
## NA's :12 NA's :18   
## behaviours\_exercise behaviours\_fruit\_veg behaviours\_less\_bad\_food  
## Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :2.000 Median :1.000 Median :1.000   
## Mean :1.585 Mean :1.082 Mean :1.254   
## 3rd Qu.:2.000 3rd Qu.:1.000 3rd Qu.:1.000   
## Max. :3.000 Max. :3.000 Max. :3.000   
## NA's :4 NA's :7 NA's :12   
## behaviours\_less\_salt behaviours\_eat\_protein behaviours\_eat\_carbs  
## Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :1.000 Median :1.000 Median :1.000   
## Mean :1.142 Mean :1.022 Mean :1.229   
## 3rd Qu.:1.000 3rd Qu.:1.000 3rd Qu.:1.000   
## Max. :3.000 Max. :3.000 Max. :3.000   
## NA's :5 NA's :16 NA's :7   
## behaviours\_lower\_blood\_pressure behaviours\_control\_diabetes  
## Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:1.000   
## Median :1.000 Median :1.000   
## Mean :1.654 Mean :1.812   
## 3rd Qu.:3.000 3rd Qu.:3.000   
## Max. :3.000 Max. :3.000   
## NA's :8 NA's :7   
## behaviours\_smoke\_less seen\_nurse\_last\_year breastfeeding   
## Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :3.000 Median :1.000 Median :1.000   
## Mean :2.473 Mean :1.024 Mean :1.011   
## 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:1.000   
## Max. :3.000 Max. :3.000 Max. :2.000   
## NA's :18 NA's :14 NA's :30   
## breastfeeding\_how\_often breastfeeding\_meals\_per\_day ante\_natal\_vitamins  
## Min. : 3.000 Min. :1.000 Min. :1.000   
## 1st Qu.: 8.000 1st Qu.:3.000 1st Qu.:1.000   
## Median :10.000 Median :3.000 Median :1.000   
## Mean : 9.328 Mean :3.221 Mean :1.043   
## 3rd Qu.:10.000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :32.000 Max. :7.000 Max. :2.000   
## NA's :29 NA's :10 NA's :184   
## pregnancy\_wanted pregnancy\_wanted\_why  
## Min. :1.000 :961   
## 1st Qu.:1.000 beause of financial problems : 1   
## Median :1.000 i did the injection and it failed me: 1   
## Mean :1.041 injectable failed me. : 1   
## 3rd Qu.:1.000 it failed : 1   
## Max. :2.000 the guy did not agree : 1   
## NA's :820   
## family\_planning\_important family\_planning\_important\_yes\_why  
## Min. :1.000 :825   
## 1st Qu.:1.000 to space up birth : 40   
## Median :1.000 to avoid unwanted pregnancy : 13   
## Mean :1.065 to prevent unwanted pregnancy: 12   
## 3rd Qu.:1.000 space up birth : 5   
## Max. :2.000 to avoid unwanted pregnancies: 5   
## NA's :812 (Other) : 66   
## family\_planning\_important\_no\_why used\_contraceptives  
## :956 Min. :1.000   
## it fails : 3 1st Qu.:2.000   
## because it fails : 1 Median :2.000   
## because of it side effect : 1 Mean :1.812   
## because you can go by your calender: 1 3rd Qu.:2.000   
## it's not necessary : 1 Max. :2.000   
## (Other) : 3 NA's :812   
## contraceptives\_helped mother\_hb height bp\_ratio   
## :939 Min. :0.0000 Min. :1.200 Min. : 0.100   
## no : 5 1st Qu.:1.0000 1st Qu.:1.570 1st Qu.: 1.500   
## yes: 22 Median :1.0000 Median :1.620 Median : 1.600   
## Mean :0.7922 Mean :1.626 Mean : 1.665   
## 3rd Qu.:1.0000 3rd Qu.:1.680 3rd Qu.: 1.800   
## Max. :1.0000 Max. :1.910 Max. :10.000   
## NA's :18 NA's :17 NA's :35   
## bmi delivery\_complications delivery\_facts  
## Min. :14.90 Min. :1.000 :748   
## 1st Qu.:24.80 1st Qu.:1.000 RETAINED PLACENTA : 9   
## Median :28.06 Median :1.000 PPH : 8   
## Mean :28.77 Mean :1.412 FOETAL DISTRESS : 7   
## 3rd Qu.:32.00 3rd Qu.:2.000 PIH : 6   
## Max. :58.60 Max. :2.000 EPISIOTOMY SUTURED: 5   
## NA's :61 NA's :694 (Other) :183   
## family\_planning family\_planning\_where sex\_education  
## Min. :1.000 :953 Min. :1   
## 1st Qu.:2.000 mamprobi : 4 1st Qu.:1   
## Median :2.000 korle bu : 2 Median :1   
## Mean :1.797 because it does not help me: 1 Mean :1   
## 3rd Qu.:2.000 kaneshie polyclinic : 1 3rd Qu.:1   
## Max. :2.000 korle-bu : 1 Max. :1   
## NA's :907 (Other) : 4 NA's :818   
## sex\_education\_when family\_planning\_who\_responsible  
## :820 Min. :1.000   
## on radio : 33 1st Qu.:3.000   
## 2008 : 19 Median :3.000   
## on television: 15 Mean :2.904   
## 2001 : 7 3rd Qu.:3.000   
## 2006 : 5 Max. :3.000   
## (Other) : 67 NA's :820   
## family\_planning\_who\_responsible\_specific family\_planning\_charges  
## Mode:logical Min. :1.000   
## NA's:966 1st Qu.:1.000   
## Median :2.000   
## Mean :1.567   
## 3rd Qu.:2.000   
## Max. :2.000   
## NA's :936   
## family\_planning\_charges\_expensive  
## Min. :1.000   
## 1st Qu.:2.000   
## Median :2.000   
## Mean :1.923   
## 3rd Qu.:2.000   
## Max. :2.000   
## NA's :953

# The summary above suggests that many of the variables are interpreted   
# as numeric when instead they should be interpreted as factors (i.e. categorical)  
# To do this:  
  
tmp <- c(  
 "health\_centre",  
 "health\_status",  
 "health\_insurance\_status",  
 "parity",  
 "works",  
 "occupation\_main",  
 "religion",  
 "belief\_cannot\_control\_health",  
 "belief\_things\_do\_affect\_pregnancy",  
 "behaviours\_exercise",  
 "behaviours\_fruit\_veg",  
 "behaviours\_less\_bad\_food",  
 "behaviours\_less\_salt",  
 "behaviours\_eat\_protein",  
 "behaviours\_eat\_carbs",  
 "behaviours\_lower\_blood\_pressure",  
 "behaviours\_control\_diabetes",  
 "behaviours\_smoke\_less",  
 "seen\_nurse\_last\_year",  
 "breastfeeding",  
 "ante\_natal\_vitamins",  
 "pregnancy\_wanted",  
 "delivery\_complications",  
 "family\_planning\_important",  
 "used\_contraceptives",  
 "family\_planning",  
 "family\_planning\_who\_responsible",  
 "family\_planning\_charges",  
 "family\_planning\_charges\_expensive"  
 )  
  
for (i in 1:length(tmp)){  
 data\_main[[tmp[i]]] <- as.factor(data\_main[[tmp[i]]])  
}  
rm(tmp)  
  
# Now to summarise again  
  
summary(data\_main)

## X record\_number hosp\_number health\_centre  
## Min. : 1.0 Min. : 1.0 : 12 1 :696   
## 1st Qu.:242.2 1st Qu.: 246.8 1005/08 : 2 2 :265   
## Median :483.5 Median : 570.5 1006/08 : 2 NA's: 5   
## Mean :483.5 Mean : 542.8 1083/08 : 2   
## 3rd Qu.:724.8 3rd Qu.: 814.2 1084/09 : 2   
## Max. :966.0 Max. :1064.0 11027/08: 2   
## NA's :2 (Other) :944   
## hospital\_visited num\_hospital\_visits health\_status age   
## Min. :1.000 Min. : 1.000 1 :375 Min. :1.000   
## 1st Qu.:2.000 1st Qu.: 4.000 2 :430 1st Qu.:4.000   
## Median :2.000 Median : 6.000 3 :115 Median :4.000   
## Mean :1.944 Mean : 6.309 4 : 21 Mean :4.427   
## 3rd Qu.:2.000 3rd Qu.: 8.000 NA's: 25 3rd Qu.:5.000   
## Max. :2.000 Max. :16.000 Max. :8.000   
## NA's :8 NA's :35 NA's :24   
## health\_insurance\_status education marital\_status occupation   
## 1 :947 Min. :1.000 Min. :1.00 BUSINESSMAN: 92   
## 2 : 9 1st Qu.:2.000 1st Qu.:2.00 DRIVER : 92   
## NA's: 10 Median :3.000 Median :2.00 TRADER : 65   
## Mean :3.124 Mean :1.88 : 47   
## 3rd Qu.:4.000 3rd Qu.:2.00 MASON : 35   
## Max. :5.000 Max. :5.00 MECHANIC : 24   
## NA's :14 NA's :26 (Other) :611   
## parity household\_size works occupation\_main  
## 1 :417 Min. : 1.000 1 :310 1 : 13   
## 2 :261 1st Qu.: 3.000 2 :635 2 : 12   
## 3 :175 Median : 4.000 NA's: 21 3 :102   
## 4 : 64 Mean : 4.179 NA's:839   
## 5 : 30 3rd Qu.: 5.000   
## (Other): 15 Max. :10.000   
## NA's : 4 NA's :18   
## occupation\_main\_specific monthly\_income religion   
## :321 Min. :1.0 1 :113   
## TRADER :215 1st Qu.:1.0 2 :821   
## SEAMSTRESS : 77 Median :1.0 3 : 6   
## HAIRDRESSER: 48 Mean :1.6 4 : 7   
## trader : 40 3rd Qu.:2.0 NA's: 19   
## CATERER : 19 Max. :4.0   
## (Other) :246 NA's :314   
## belief\_cannot\_control\_health belief\_things\_do\_affect\_pregnancy  
## 1 :437 1 :356   
## 2 :229 2 :138   
## 3 :127 3 :148   
## 4 :133 4 :242   
## 5 : 28 5 : 64   
## NA's: 12 NA's: 18   
##   
## behaviours\_exercise behaviours\_fruit\_veg behaviours\_less\_bad\_food  
## 1 :415 1 :888 1 :765   
## 2 :531 2 : 63 2 :136   
## 3 : 16 3 : 8 3 : 53   
## NA's: 4 NA's: 7 NA's: 12   
##   
##   
##   
## behaviours\_less\_salt behaviours\_eat\_protein behaviours\_eat\_carbs  
## 1 :856 1 :931 1 :839   
## 2 : 74 2 : 17 2 : 20   
## 3 : 31 3 : 2 3 :100   
## NA's: 5 NA's: 16 NA's: 7   
##   
##   
##   
## behaviours\_lower\_blood\_pressure behaviours\_control\_diabetes  
## 1 :594 1 :515   
## 2 :101 2 :109   
## 3 :263 3 :335   
## NA's: 8 NA's: 7   
##   
##   
##   
## behaviours\_smoke\_less seen\_nurse\_last\_year breastfeeding  
## 1 :214 1 :938 1 :926   
## 2 : 72 2 : 5 2 : 10   
## 3 :662 3 : 9 NA's: 30   
## NA's: 18 NA's: 14   
##   
##   
##   
## breastfeeding\_how\_often breastfeeding\_meals\_per\_day ante\_natal\_vitamins  
## Min. : 3.000 Min. :1.000 1 :748   
## 1st Qu.: 8.000 1st Qu.:3.000 2 : 34   
## Median :10.000 Median :3.000 NA's:184   
## Mean : 9.328 Mean :3.221   
## 3rd Qu.:10.000 3rd Qu.:3.000   
## Max. :32.000 Max. :7.000   
## NA's :29 NA's :10   
## pregnancy\_wanted pregnancy\_wanted\_why  
## 1 :140 :961   
## 2 : 6 beause of financial problems : 1   
## NA's:820 i did the injection and it failed me: 1   
## injectable failed me. : 1   
## it failed : 1   
## the guy did not agree : 1   
##   
## family\_planning\_important family\_planning\_important\_yes\_why  
## 1 :144 :825   
## 2 : 10 to space up birth : 40   
## NA's:812 to avoid unwanted pregnancy : 13   
## to prevent unwanted pregnancy: 12   
## space up birth : 5   
## to avoid unwanted pregnancies: 5   
## (Other) : 66   
## family\_planning\_important\_no\_why used\_contraceptives  
## :956 1 : 29   
## it fails : 3 2 :125   
## because it fails : 1 NA's:812   
## because of it side effect : 1   
## because you can go by your calender: 1   
## it's not necessary : 1   
## (Other) : 3   
## contraceptives\_helped mother\_hb height bp\_ratio   
## :939 Min. :0.0000 Min. :1.200 Min. : 0.100   
## no : 5 1st Qu.:1.0000 1st Qu.:1.570 1st Qu.: 1.500   
## yes: 22 Median :1.0000 Median :1.620 Median : 1.600   
## Mean :0.7922 Mean :1.626 Mean : 1.665   
## 3rd Qu.:1.0000 3rd Qu.:1.680 3rd Qu.: 1.800   
## Max. :1.0000 Max. :1.910 Max. :10.000   
## NA's :18 NA's :17 NA's :35   
## bmi delivery\_complications delivery\_facts  
## Min. :14.90 1 :160 :748   
## 1st Qu.:24.80 2 :112 RETAINED PLACENTA : 9   
## Median :28.06 NA's:694 PPH : 8   
## Mean :28.77 FOETAL DISTRESS : 7   
## 3rd Qu.:32.00 PIH : 6   
## Max. :58.60 EPISIOTOMY SUTURED: 5   
## NA's :61 (Other) :183   
## family\_planning family\_planning\_where sex\_education  
## 1 : 12 :953 Min. :1   
## 2 : 47 mamprobi : 4 1st Qu.:1   
## NA's:907 korle bu : 2 Median :1   
## because it does not help me: 1 Mean :1   
## kaneshie polyclinic : 1 3rd Qu.:1   
## korle-bu : 1 Max. :1   
## (Other) : 4 NA's :818   
## sex\_education\_when family\_planning\_who\_responsible  
## :820 1 : 7   
## on radio : 33 3 :139   
## 2008 : 19 NA's:820   
## on television: 15   
## 2001 : 7   
## 2006 : 5   
## (Other) : 67   
## family\_planning\_who\_responsible\_specific family\_planning\_charges  
## Mode:logical 1 : 13   
## NA's:966 2 : 17   
## NA's:936   
##   
##   
##   
##   
## family\_planning\_charges\_expensive  
## 1 : 1   
## 2 : 12   
## NA's:953   
##   
##   
##   
##

###  
  
qplot(bmi, height, data=data\_main)

## Warning: Removed 61 rows containing missing values (geom\_point).



# Suggests just one possible error now, a height under 1.3

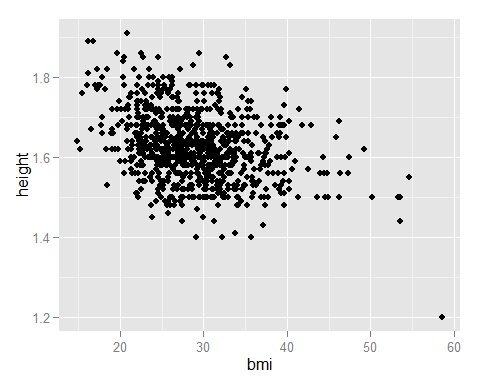
Creating an error data frame, containing record numbers and reasons why the record numbers appear to have issues.

error\_df <- data.frame(  
 record\_number=subset(data\_main, height < 1.3)$record\_number,  
 reason="outlier: low height"  
 )  
  
###

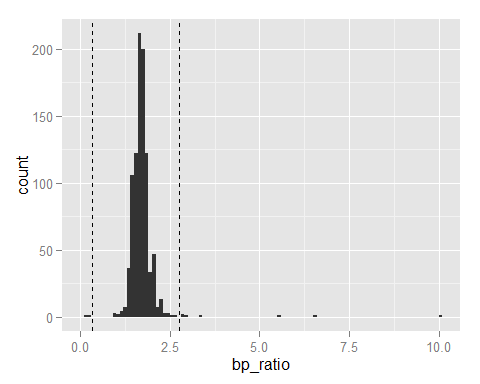
Exploring data and identifying further potential errors

qplot(bmi, height, data=data\_main)

## Warning: Removed 61 rows containing missing values (geom\_point).

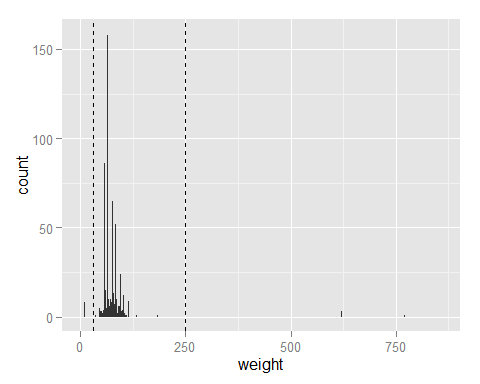


# Suggests just one possible error now, a height under 1.3  
  
  
error\_df <- data.frame(  
 record\_number=subset(data\_main, height < 1.3)$record\_number,  
 reason="outlier: low height"  
 )  
  
# Exploring any potential outliers in bp\_ratio  
  
qplot(bp\_ratio, data=data\_main, binwidth=0.1) + geom\_vline(aes(xintercept=c(0.32, 2.75)), linetype="dashed")



# adding additional records to error\_df  
  
error\_df <- rbind(  
 error\_df,  
 data.frame(  
 record\_number=subset(data\_main, bp\_ratio < 0.32)$record\_number,  
 reason="bp ratio almost 0"  
 ),  
 data.frame(  
 record\_number=subset(data\_main, bp\_ratio>2.75)$record\_number,  
 reason="bp ratio over 2.75"  
 )  
 )  
  
  
# Exploring weights  
# Very high weights:  
qplot(weight, data=data\_visits, binwidth=0.2) + geom\_vline(aes(xintercept=c(30, 250)), linetype="dashed")

## Warning: position\_stack requires constant width: output may be incorrect

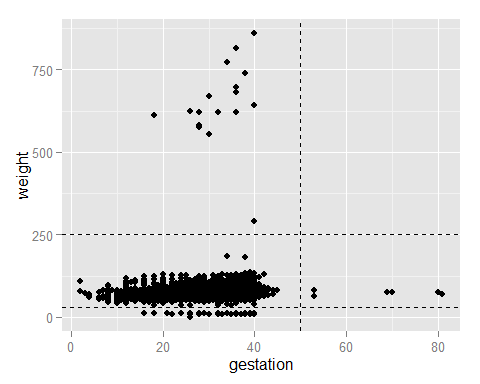


# Adding records to error\_df  
error\_df <- rbind(  
 error\_df,  
 data.frame(  
 record\_number=subset(data\_visits, weight < 30)$record\_number,  
 reason="weight below 30"  
 ),  
 data.frame(  
 record\_number=subset(data\_visits, weight > 250)$record\_number,  
 reason="weight above 250"  
 )  
 )  
  
  
## now to look at the visits data  
  
data\_visits

## Source: local data frame [7,728 x 10]  
##   
## X record\_number visit\_number date weight bp\_num bp\_denom  
## 1 1 1 1 09-Jul-08 51.0 100 70  
## 2 2 2 1 NA 66.0 100 60  
## 3 3 3 1 09-Dec-08 56.5 110 70  
## 4 4 4 1 09-Sep-08 77.0 110 60  
## 5 5 5 1 10-Sep-08 57.0 110 80  
## 6 6 6 1 06-Jun-08 83.0 120 60  
## 7 7 7 1 10-Jul-08 82.0 110 70  
## 8 8 8 1 24-Aug-08 70.0 130 90  
## 9 9 9 1 20-Oct-08 58.0 100 60  
## 10 10 10 1 14-Feb-08 59.0 100 60  
## .. .. ... ... ... ... ... ...  
## Variables not shown: urine\_protein (fctr), urine\_sugar (fctr), gestation  
## (int)

qplot(y=weight, x=gestation, data=data\_visits) + geom\_vline(aes(xintercept=c(50)), linetype="dashed") +   
 geom\_hline(aes(yintercept=c(30,250)), linetype="dashed")

## Warning: Removed 3123 rows containing missing values (geom\_point).



error\_df <- rbind(  
 error\_df,  
 data.frame(  
 record\_number=subset(data\_visits, gestation > 50)$record\_number,  
 reason="gestation over 50"  
 )  
)  
  
  
  
# now to try lubridate  
  
dates\_of\_interest <- data\_visits$date  
  
# need to change months into values   
  
  
replace\_runner <- function(target, pattern, replace){   
 n <- length(pattern)  
 for (i in 1:n){  
 target <- str\_replace\_all(target, pattern=pattern[i], replace=replace[i])  
 }   
 return(target)  
}  
   
dates\_of\_interest <- replace\_runner(  
 dates\_of\_interest,  
 pattern=c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"),  
 replace=c("01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11", "12")  
 )  
  
  
# now to use lubridate, specifying day, month, year format  
  
dates\_of\_interest <- dmy(dates\_of\_interest)  
data\_visits$date <- dates\_of\_interest  
  
# can now calculate difference in number of days and so on.   
  
# want to calculate dates since first visit  
  
fn <- function(x){  
 t0 <- x$date[x$visit\_number==1]  
 dates <- x$date  
 n <- length(dates)  
 days\_since\_first\_visit <- rep(NA, n)  
 for (i in 1:n){   
 tmp <- dates[i] - t0  
 days\_since\_first\_visit[i] <- as.numeric(tmp)  
 }  
 out <- data.frame(x, days\_since\_first\_visit=days\_since\_first\_visit)  
 return(out)  
}  
  
  
data\_visits <- ddply(data\_visits, .(record\_number), fn)

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

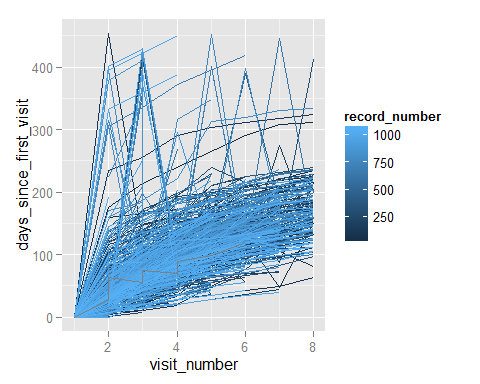
## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

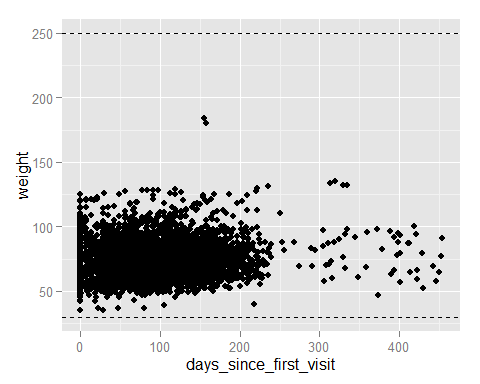
## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

## Warning in days\_since\_first\_visit[i] <- as.numeric(tmp): number of items  
## to replace is not a multiple of replacement length

# process  
# 1) remove NA entries to visit number  
# 2) remove negative days\_since\_first\_visit  
# 3) remove positive days\_since\_first\_visit greater than top 1% of values  
  
data\_visits <- data\_visits %>% tbl\_df()  
  
data\_visits\_ok <- data\_visits %>%   
 filter(!is.na(date)) %>%  
 filter(days\_since\_first\_visit >= 0) %>%   
 filter(days\_since\_first\_visit < quantile(data\_visits$days\_since\_first\_visit, 0.99, na.rm=T)) %>%  
 filter(weight > 30 & weight < 250) %>%  
 filter(gestation < 50)  
  
# Plot days since first visit and visit number   
# for the records still retained  
  
  
qplot(  
 x=visit\_number,  
 y=days\_since\_first\_visit,   
 group=record\_number,  
 colour=record\_number,  
 data=data\_visits\_ok,  
 geom="line"  
 )



qplot(  
 x=days\_since\_first\_visit,  
 y=weight,   
 data=data\_visits\_ok  
) + geom\_hline(mapping=aes(yintercept=c(30,250)), linetype="dashed")



#Saving file with records that seem OK  
# (Commenting out as file already created)  
  
# write.csv(  
# data\_visits\_ok,  
# file="data/data\_visits\_with\_days\_since\_first\_visit\_ok\_only.csv",  
# row.names=F  
# )  
  
# Take a random sample of 25 records and display change in weight over time  
  
tmp <- data\_visits\_ok %>% select(record\_number) %>% unique() %>% sample\_n(25) %>% unlist() %>% as.numeric()  
  
data\_visits\_ok %>% filter(record\_number %in% tmp) %>% ggplot() +  
 geom\_line(aes(x=visit\_number, y=weight)) +  
 facet\_wrap( ~ record\_number)

## geom\_path: Each group consist of only one observation. Do you need to adjust the group aesthetic?

