

# AmericanRadiation

April 29, 2018

```
In [156]: library(dplyr)
          library(stringr)
          library(tidyr)
          library(cowplot)
          options(digits=10)
          setwd("../Data")

          ##read all the downloaded files
          myfiles <- list.files(pattern="*.csv")
          files <- do.call(rbind, lapply(myfiles, function(x) read.csv(x, stringsAsFactors = FALSE)))
```

```
In [157]: head(files)
```

MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
942	AK: FAIRBANKS	02/24/2009 22:29:00	NA
942	AK: FAIRBANKS	02/24/2009 22:40:00	NA
942	AK: FAIRBANKS	02/24/2009 23:53:00	NA
942	AK: FAIRBANKS	02/25/2009 00:54:00	NA
942	AK: FAIRBANKS	02/25/2009 01:54:00	NA
942	AK: FAIRBANKS	02/25/2009 02:54:00	NA

```
In [158]: #check if files are loaded correctly
          table(files$LOCATION_NAME)
```

AK: FAIRBANKS	CA: SAN FRANCISCO	CO: DENVER	DC: WASHINGTON
52233	93358	89195	91669
FL: MIAMI	HI: HONOLULU	IA: DES MOINES	MN: DULUTH
66205	76741	82205	72126
NY: NEW YORK CITY	TX: HOUSTON	WA: SEATTLE	
75573	79544	77690	

```
In [159]: length(files$MONITOR.ID)
```

856539

```
In [160]: #split the sample collection time into time, day, month, year
          files["month"] <- substr(files$SAMPLE.COLLECTION.TIME, 1, 2)
```

```
files["day"] <- substr(files$SAMPLE.COLLECTION.TIME, 4, 5)
files["year"] <- substr(files$SAMPLE.COLLECTION.TIME, 7, 10)
files["time"] <- substr(files$SAMPLE.COLLECTION.TIME, 12, 13)
head(files)
```

MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
942	AK: FAIRBANKS	02/24/2009 22:29:00	NA
942	AK: FAIRBANKS	02/24/2009 22:40:00	NA
942	AK: FAIRBANKS	02/24/2009 23:53:00	NA
942	AK: FAIRBANKS	02/25/2009 00:54:00	NA
942	AK: FAIRBANKS	02/25/2009 01:54:00	NA
942	AK: FAIRBANKS	02/25/2009 02:54:00	NA

In [161]: *#check if the string split did it correctly*

```
table(files$month)
table(files$day)
table(files$year)
table(files$time)
```

```
01    02    03    04    05    06    07    08    09    10    11    12
71499 67533 77264 75479 72312 67908 69584 72366 68693 68393 71175 74333
```

```
01    02    03    04    05    06    07    08    09    10    11    12    13
28000 28126 28027 27977 27952 28033 28175 28105 28279 28139 28103 28278 28057
14    15    16    17    18    19    20    21    22    23    24    25    26
28249 28210 28446 28364 28329 27703 27972 28266 28343 28283 28442 28369 28228
27    28    29    30    31
28289 28107 25962 25452 16274
```

```
2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
5643 35115 54819 60080 70296 84658 85813 89022 83261 87774 87265 85325 27468
```

```
00    01    02    03    04    05    06    07    08    09    10    11    12
35757 35827 35604 35730 35720 35679 35712 35765 35765 35818 35726 35876 35736
13    14    15    16    17    18    19    20    21    22    23
35690 35644 35676 35484 35645 35579 35605 35543 35612 35693 35653
```

In [162]: *#create a new column with month and year information*

```
files["month_year"] <- paste(files$year, files$month, sep="/")
head(files)
```

MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
942	AK: FAIRBANKS	02/24/2009 22:29:00	NA
942	AK: FAIRBANKS	02/24/2009 22:40:00	NA
942	AK: FAIRBANKS	02/24/2009 23:53:00	NA
942	AK: FAIRBANKS	02/25/2009 00:54:00	NA
942	AK: FAIRBANKS	02/25/2009 01:54:00	NA
942	AK: FAIRBANKS	02/25/2009 02:54:00	NA

In [163]: *#check the data*

```
summary(files$GAMMA.COUNT.RATE.R02..CPM.)
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  0    1197    1726    1883    2398  501564  2353
```

In [164]: `names(files)`

```
1. 'MONITOR.ID' 2. 'LOCATION_NAME' 3. 'SAMPLE.COLLECTION.TIME'
4. 'DOSE.EQUIVALENT.RATE..nSv.h.' 5. 'GAMMA.COUNT.RATE.R02..CPM.'
6. 'GAMMA.COUNT.RATE.R03..CPM.' 7. 'GAMMA.COUNT.RATE.R04..CPM.'
8. 'GAMMA.COUNT.RATE.R05..CPM.' 9. 'GAMMA.COUNT.RATE.R06..CPM.'
10. 'GAMMA.COUNT.RATE.R07..CPM.' 11. 'GAMMA.COUNT.RATE.R08..CPM.'
12. 'GAMMA.COUNT.RATE.R09..CPM.' 13. 'STATUS' 14. 'month' 15. 'day' 16. 'year' 17. 'time'
18. 'month_year'
```

In [165]: *#remove the rows where radiation values contain NA*

```
files.com <- files[complete.cases(files[, 5:12]), ]
```

In [166]: `length(files$MONITOR.ID)`

```
length(files.com$MONITOR.ID)
```

```
856539
```

```
854059
```

In [167]: `names(files.com)`

```
1. 'MONITOR.ID' 2. 'LOCATION_NAME' 3. 'SAMPLE.COLLECTION.TIME'
4. 'DOSE.EQUIVALENT.RATE..nSv.h.' 5. 'GAMMA.COUNT.RATE.R02..CPM.'
6. 'GAMMA.COUNT.RATE.R03..CPM.' 7. 'GAMMA.COUNT.RATE.R04..CPM.'
8. 'GAMMA.COUNT.RATE.R05..CPM.' 9. 'GAMMA.COUNT.RATE.R06..CPM.'
10. 'GAMMA.COUNT.RATE.R07..CPM.' 11. 'GAMMA.COUNT.RATE.R08..CPM.'
12. 'GAMMA.COUNT.RATE.R09..CPM.' 13. 'STATUS' 14. 'month' 15. 'day' 16. 'year' 17. 'time'
18. 'month_year'
```

In [168]: `head(files.com)`

MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
942	AK: FAIRBANKS	02/24/2009 22:29:00	NA
942	AK: FAIRBANKS	02/24/2009 22:40:00	NA
942	AK: FAIRBANKS	02/24/2009 23:53:00	NA
942	AK: FAIRBANKS	02/25/2009 00:54:00	NA
942	AK: FAIRBANKS	02/25/2009 01:54:00	NA
942	AK: FAIRBANKS	02/25/2009 02:54:00	NA

```
In [169]: table(files.com$LOCATION_NAME)
```

```

      AK: FAIRBANKS CA: SAN FRANCISCO      CO: DENVER      DC: WASHINGTON
      51490          93246          89098          91597
      FL: MIAMI      HI: HONOLULU      IA: DES MOINES      MN: DULUTH
      66078          76630          81996          72053
NY: NEW YORK CITY      TX: HOUSTON      WA: SEATTLE
      74788          79463          77620

```

```
In [170]: #change location name for convenient usage in the future
files.com["location"] <- substr(files.com$LOCATION_NAME, 5, 1000000L)
head(files.com)
```

MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
942	AK: FAIRBANKS	02/24/2009 22:29:00	NA
942	AK: FAIRBANKS	02/24/2009 22:40:00	NA
942	AK: FAIRBANKS	02/24/2009 23:53:00	NA
942	AK: FAIRBANKS	02/25/2009 00:54:00	NA
942	AK: FAIRBANKS	02/25/2009 01:54:00	NA
942	AK: FAIRBANKS	02/25/2009 02:54:00	NA

```
In [171]: table(files.com$location)
```

```

      DENVER      DES MOINES      DULUTH      FAIRBANKS      HONOLULU
      89098          81996          72053          51490          76630
      HOUSTON          MIAMI NEW YORK CITY SAN FRANCISCO      SEATTLE
      79463          66078          74788          93246          77620
WASHINGTON
      91597

```

```
In [172]: #look at the example in Denver, as it is the closest city near where I live
Denver <- filter(files.com, location == "DENVER")
head(Denver)
```

MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
801	CO: DENVER	08/20/2006 19:14:00	NA
801	CO: DENVER	08/20/2006 19:34:00	NA
801	CO: DENVER	08/21/2006 13:45:00	NA
801	CO: DENVER	09/26/2006 19:52:00	NA
801	CO: DENVER	09/26/2006 20:53:00	NA
801	CO: DENVER	09/26/2006 21:53:00	NA

```
In [174]: names(Denver)
```

1. 'MONITOR.ID'
2. 'LOCATION\_NAME'
3. 'SAMPLE.COLLECTION.TIME'
4. 'DOSE.EQUIVALENT.RATE..nSv.h.'
5. 'GAMMA.COUNT.RATE.R02..CPM.'

```

6. 'GAMMA.COUNT.RATE.R03..CPM.' 7. 'GAMMA.COUNT.RATE.R04..CPM.'
8. 'GAMMA.COUNT.RATE.R05..CPM.' 9. 'GAMMA.COUNT.RATE.R06..CPM.'
10. 'GAMMA.COUNT.RATE.R07..CPM.' 11. 'GAMMA.COUNT.RATE.R08..CPM.'
12. 'GAMMA.COUNT.RATE.R09..CPM.' 13. 'STATUS' 14. 'month' 15. 'day' 16. 'year' 17. 'time'
18. 'month_year' 19. 'location'

```

In [175]: *#shrink the data, delete unnecessary information*

```

Denver.slim <- Denver[, c(5:12,14:18)]
names(Denver.slim)[1:8] <- c("R02", "R03", "R04", "R05", "R06", "R07", "R08", "R09")

head(Denver.slim)

```

R02	R03	R04	R05	R06	R07	R08	R09	month	day	year	time	month_year
4515	610	308	288	402	221	34	37	08	20	2006	19	2006/08
4493	608	304	280	398	231	33	38	08	20	2006	19	2006/08
4552	615	312	288	398	231	35	38	08	21	2006	13	2006/08
3316	409	194	183	250	146	23	29	09	26	2006	19	2006/09
3177	416	200	190	257	153	25	30	09	26	2006	20	2006/09
3137	406	204	189	259	157	25	30	09	26	2006	21	2006/09

In [176]: *#convert from wide to long format*

```

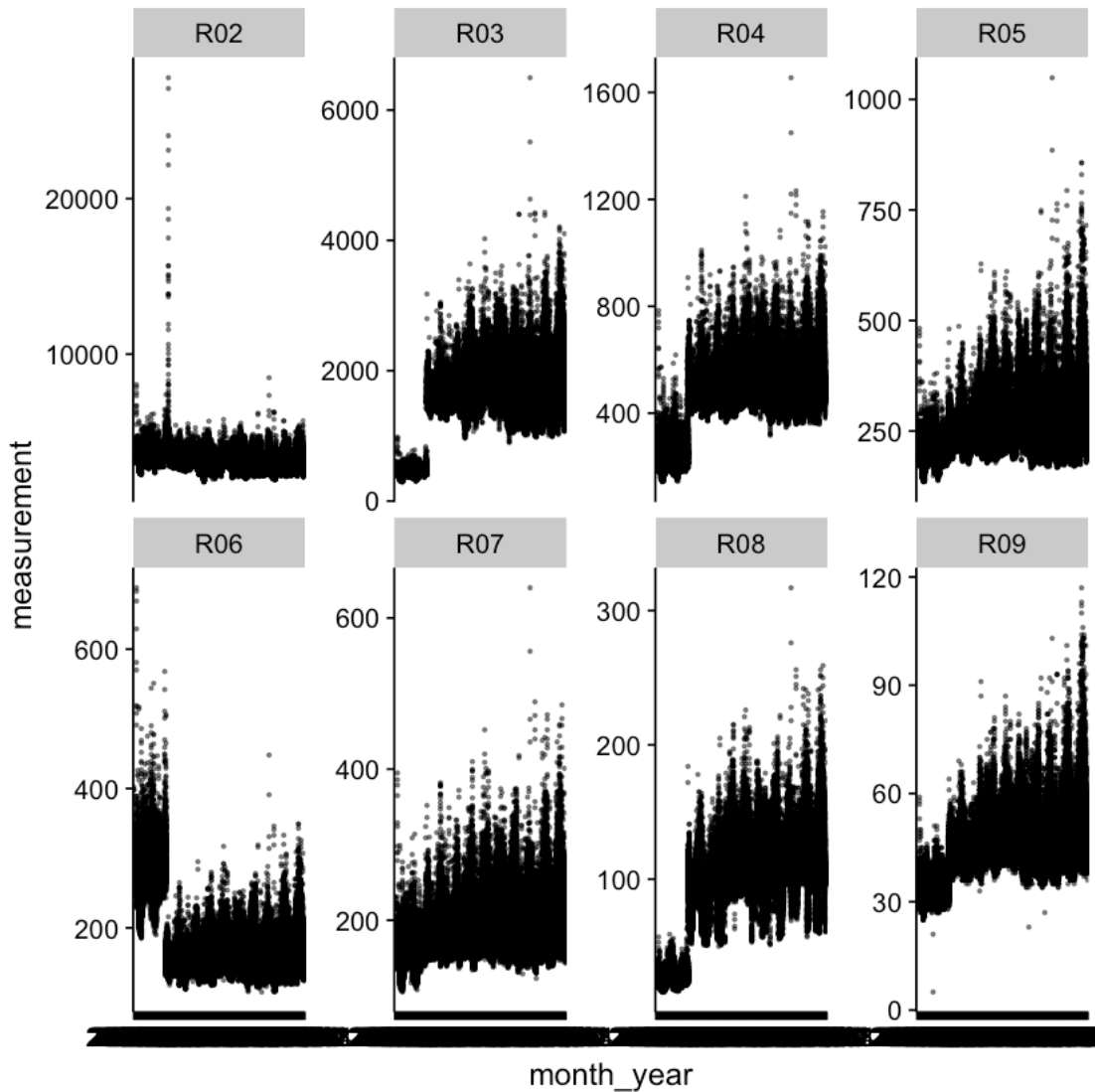
Denver.long <- gather(Denver.slim, rate, measurement, R02:R09)

```

In [177]: `head(Denver.long)`

month	day	year	time	month_year	rate	measurement
08	20	2006	19	2006/08	R02	4515
08	20	2006	19	2006/08	R02	4493
08	21	2006	13	2006/08	R02	4552
09	26	2006	19	2006/09	R02	3316
09	26	2006	20	2006/09	R02	3177
09	26	2006	21	2006/09	R02	3137

In [178]: *#plot out how each level of radiation change across months in the recent >10 years in*  
`ggplot(data=Denver.long, aes(x=month_year, y=measurement)) + geom_point(size=0.5, alph`

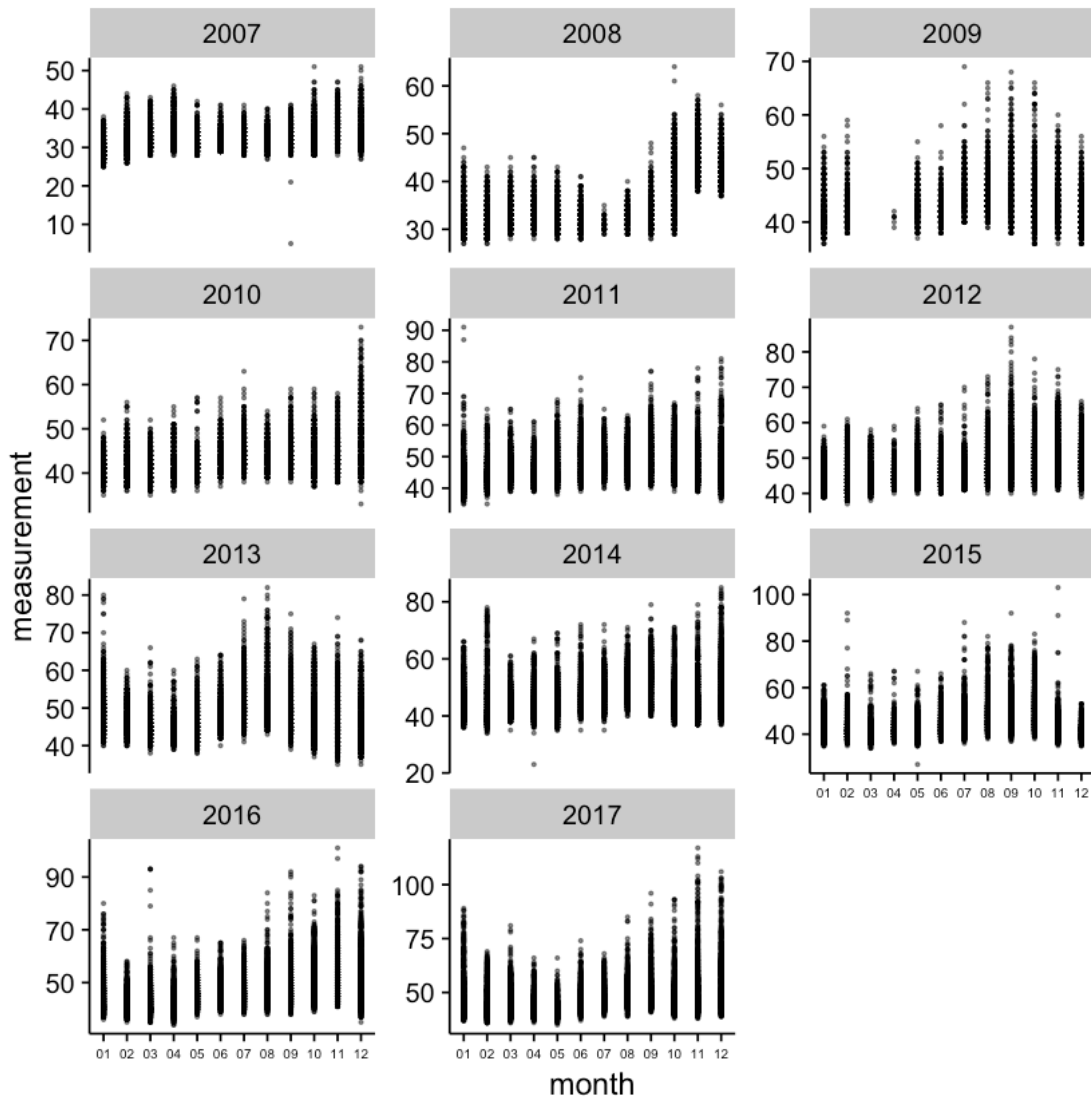


```
In [179]: #it showed when the radiation harzard increases, the radition level increase across ye
          ggsave("Denver.monthAndYearTrend.pdf")
```

Saving 7 x 7 in image

```
In [180]: #foucus on R09, to see how radiation level changes across month in each year
          Denver07_17_R09 <- filter(Denver.long, year!= "2006" & year!= "2018") %>% filter(rate=
```

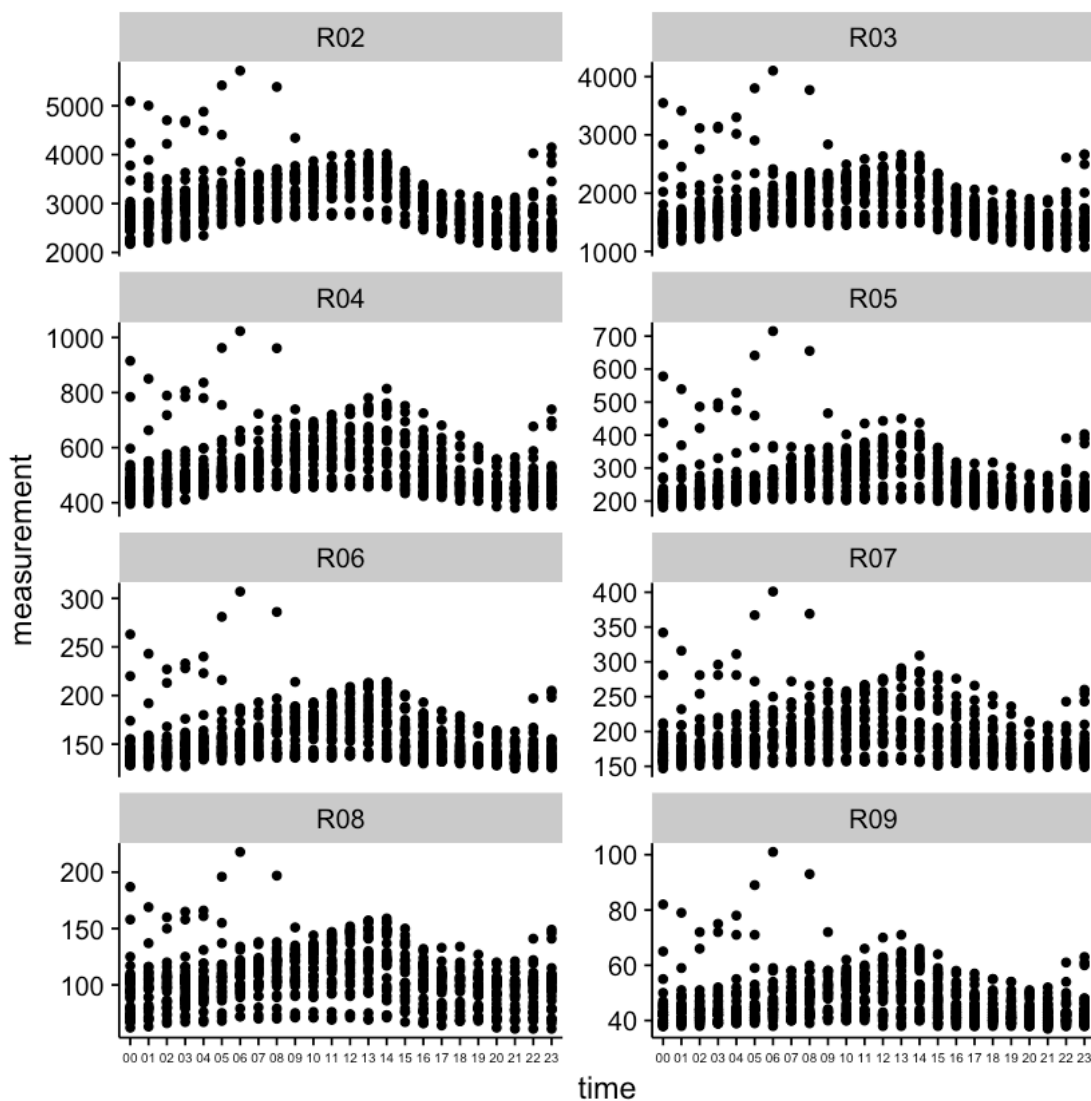
```
In [181]: ggplot(data=Denver07_17_R09, aes(x=month, y=measurement)) + facet_wrap(~year, scales="
```



```
In [182]: #no obvious pattern found, generally lower radiation in summer but higher in winter (0
          ggsave("DenverR09_month.pdf")
```

Saving 7 x 7 in image

```
In [183]: #focus on data from 2018 March in Denver to see how radiation change across time withi
          #the highest radiation appears at around 12-13pm
          Denver2018March <- filter(Denver.long, year==2018 & month=="03")
          ggplot(Denver2018March, aes(x=time, y=measurement)) + geom_smooth(method = "loess", co
```



```
In [184]: ggsave("Denver2018March_timePattern.pdf")
```

Saving 7 x 7 in image

```
In [185]: #shift to the big dataset including multiple states
names(files.com)[5:12] <- c("R02", "R03", "R04", "R05", "R06", "R07", "R08", "R09")
head(files.com)
```

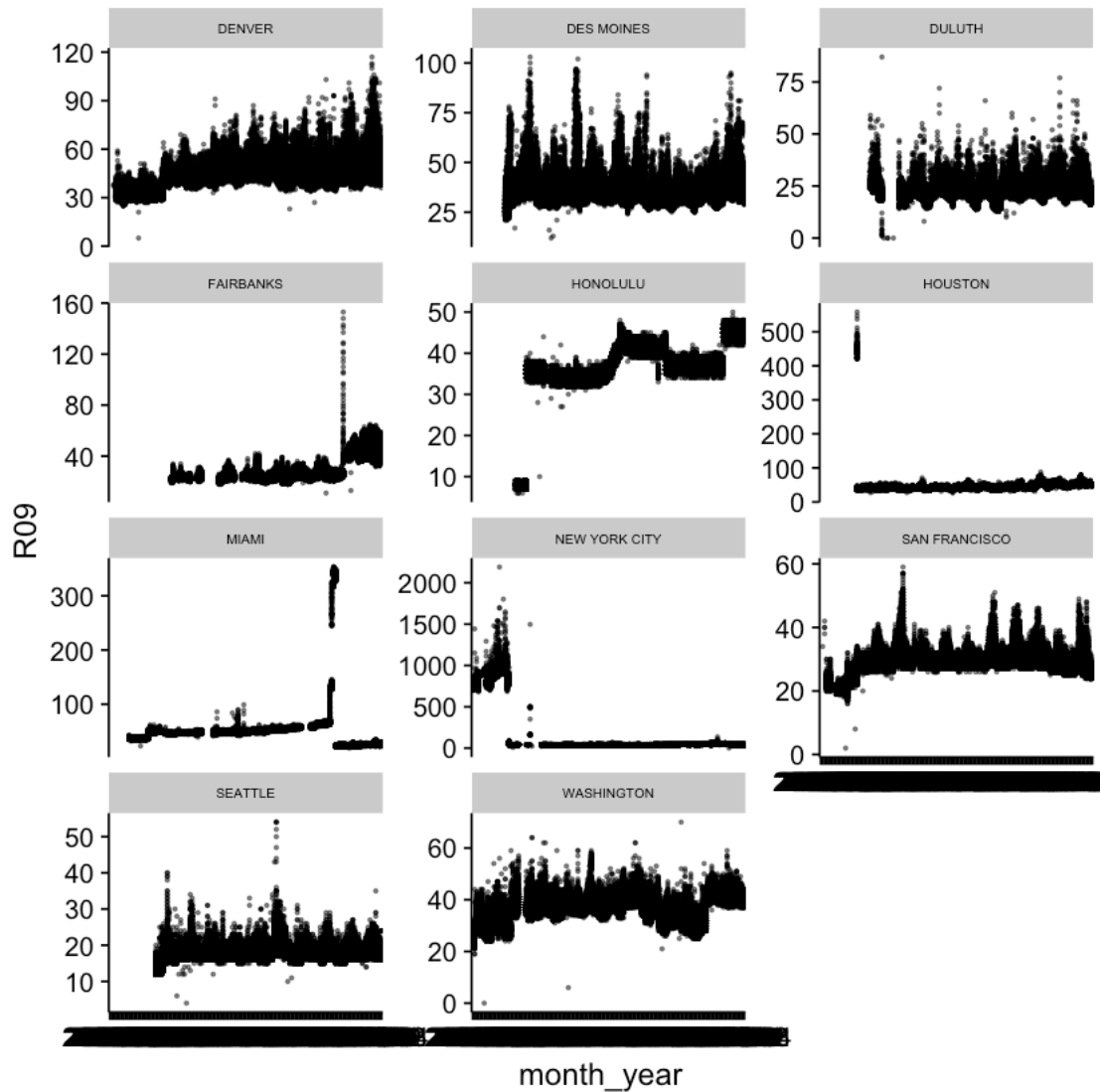


MONITOR.ID	LOCATION_NAME	SAMPLE.COLLECTION.TIME	DOSE.EQUIVALENT.RATE..nSv.h
942	AK: FAIRBANKS	02/24/2009 22:29:00	NA
942	AK: FAIRBANKS	02/24/2009 22:40:00	NA
942	AK: FAIRBANKS	02/24/2009 23:53:00	NA
942	AK: FAIRBANKS	02/25/2009 00:54:00	NA
942	AK: FAIRBANKS	02/25/2009 01:54:00	NA
942	AK: FAIRBANKS	02/25/2009 02:54:00	NA

```
In [186]: files.slim <- files.com[, c(5:12, 14:19)]
          head(files.slim)
```

R02	R03	R04	R05	R06	R07	R08	R09	month	day	year	time	month_year	location
1060	571	169	77	51	75	32	20	02	24	2009	22	2009/02	FAIRBANKS
1068	602	172	83	51	83	34	19	02	24	2009	22	2009/02	FAIRBANKS
1105	610	182	81	55	79	34	22	02	24	2009	23	2009/02	FAIRBANKS
1093	608	184	79	53	80	36	22	02	25	2009	00	2009/02	FAIRBANKS
1097	615	186	84	54	80	35	22	02	25	2009	01	2009/02	FAIRBANKS
1080	603	185	81	54	83	34	22	02	25	2009	02	2009/02	FAIRBANKS

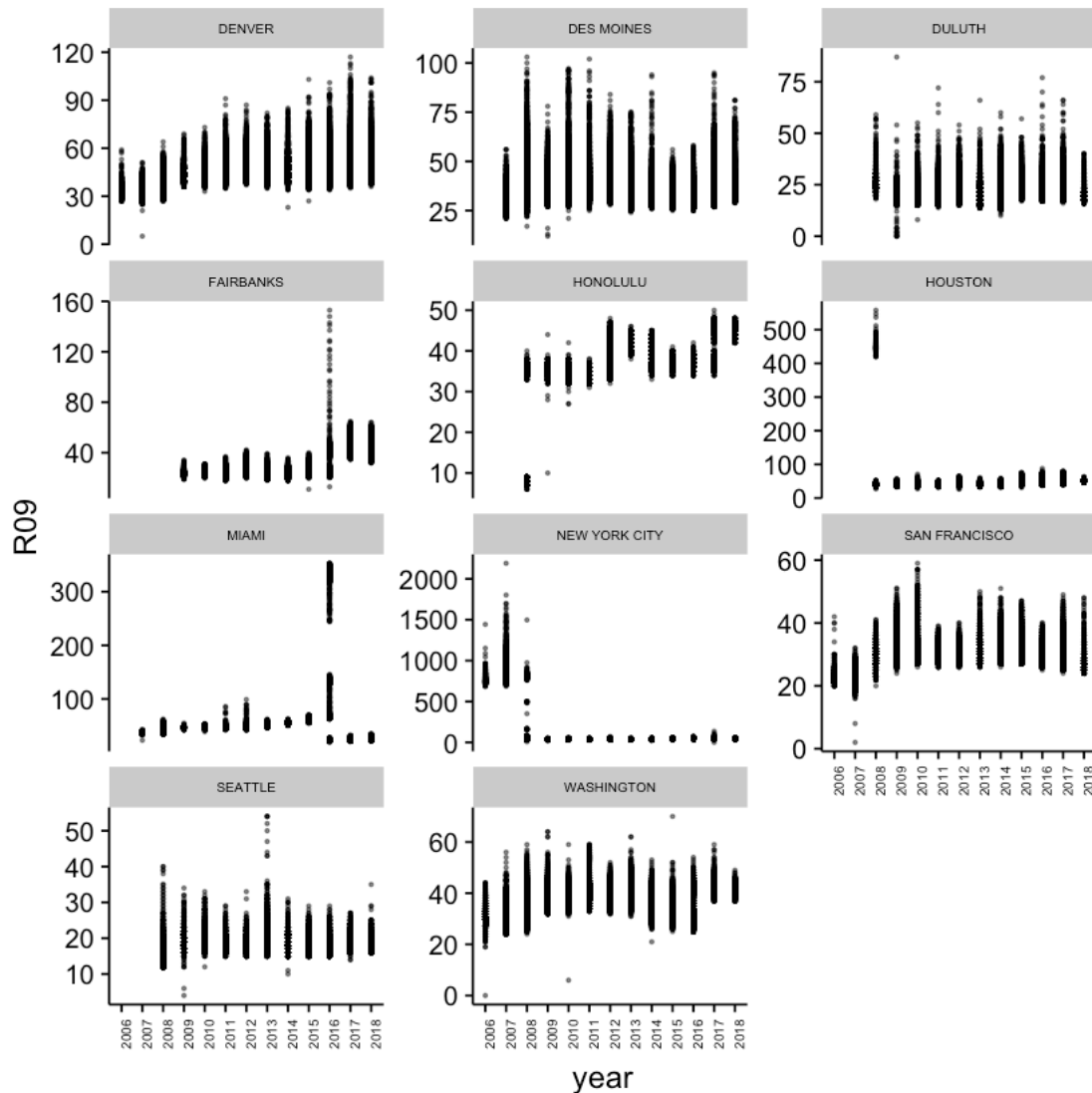
```
In [187]: # to investigate how the most harzardous radiation change across years in different st
          #and found not all states increases the radiation level across time, some decrease a l
          # further question could be: what policies has NYC taken to reduce the radiation level
          ggplot(data=files.slim, aes(x=month_year, y=R09)) + geom_point(size=0.5, alpha=0.5) +
```



```
In [188]: ggsave("representativeStateR09Pattern_month_year.pdf")
```

Saving 7 x 7 in image

```
In [189]: #indicate which year the radiation changed dramatically
ggplot(data=files.slim, aes(x=year, y=R09)) + geom_point(size=0.5, alpha=0.5) + facet_
```



```
In [190]: ggsave("representativeStateR09Pattern_year.pdf")
```

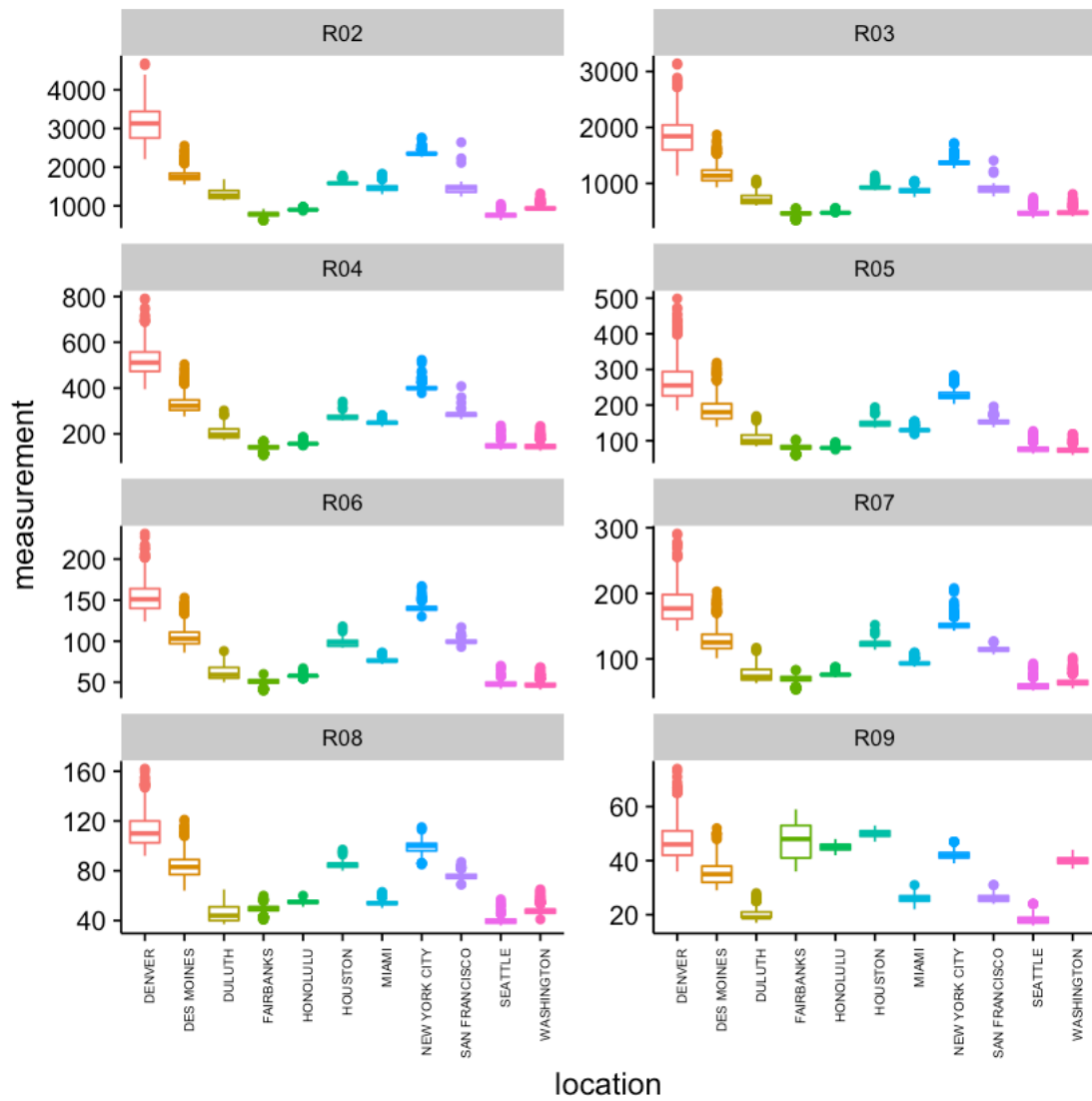
Saving 7 x 7 in image

```
In [191]: #restrict the data to 2018 April, the most recent month, to compare the radiation diff
```

```
files2018April <- filter(files.slim, year=="2018" & month=="04")
head(files2018April)
files2018April.long <- gather(files2018April, rate, measurement, R02:R09)
```

```
#Denver has the highest radiation, could it be related with the mining industry neaby?
#fairbanks, duluth, honolulu etc showed lower radiation
ggplot(data=files2018April.long, aes(x=location, y=measurement)) + geom_boxplot(aes(co
```

R02	R03	R04	R05	R06	R07	R08	R09	month	day	year	time	month_year	location
683	397	122	69	44	62	45	52	04	01	2018	00	2018/04	FAIRBANKS
661	378	119	67	44	61	43	52	04	01	2018	01	2018/04	FAIRBANKS
673	381	116	66	43	58	44	51	04	01	2018	02	2018/04	FAIRBANKS
696	389	124	68	43	61	44	53	04	01	2018	03	2018/04	FAIRBANKS
707	397	125	71	48	61	45	51	04	01	2018	04	2018/04	FAIRBANKS
695	388	122	68	45	59	45	50	04	01	2018	05	2018/04	FAIRBANKS



In [192]: `ggsave("representativeState2018AprilPattern.pdf")`

Saving 7 x 7 in image