

# Bicycle/Auto Collisions

*Alan K Jackson*

*July 14, 2018*

## Bicyclist/Automobile Incidents in Harris County

Based on the TxDOT data from 2010 to 2017, I have analyzed incidents where cyclists got hit by cars. First we will begin with some general statistics, and then bore down to some more specific items.

### CDC Data

From the website for the Centers for Disease Control, we have the following information regarding Bicyclist incidents:

Who is most at risk?

Adults aged 50 to 59 years have the highest bicycle death rates.

Children (5-14 years) and adolescents (15-19 years) have the highest rates of nonfatal bicycle-related injuries, accounting for more than one-third of all bicycle-related injuries seen in U.S. emergency departments.

Males die 6 times more often and are injured 4 times more often on bicycles than females.

Most bicyclist deaths occur in urban areas and at non-intersection locations.

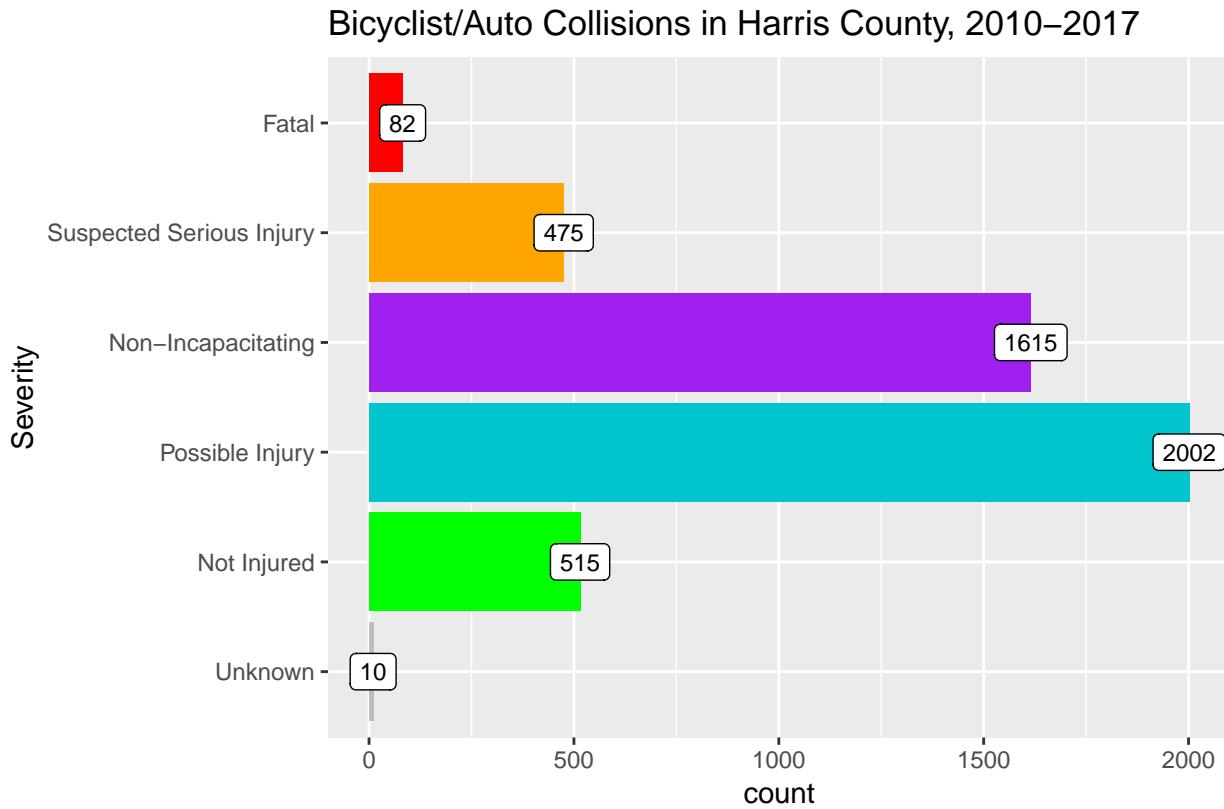
Among bicyclist deaths, 37% had alcohol involvement either for the motor vehicle driver or bicycle rider.

### Injury Distribution

The state has six categories of injury (if we include “Unknown”). Somewhat surprisingly, there are about 80 Bicyclist fatalities per year in the county. And about 2 incidents per day, on average.

It is important to keep in mind that there is almost certainly heavy reporting bias in these numbers. Every fatality and serious injury will likely get reported. Minor injuries may escape being reported, and non-injury incidents are probably not reported most of the time.

```
cyclists %>%
  ggplot(aes(x=Severity)) +
  geom_bar(fill=SevereColor) +
  geom_label(aes(label=..count..), stat='count', size = 3) +
  coord_flip() +
  labs(title = "Bicyclist/Auto Collisions in Harris County, 2010-2017",
       caption = caption)
```



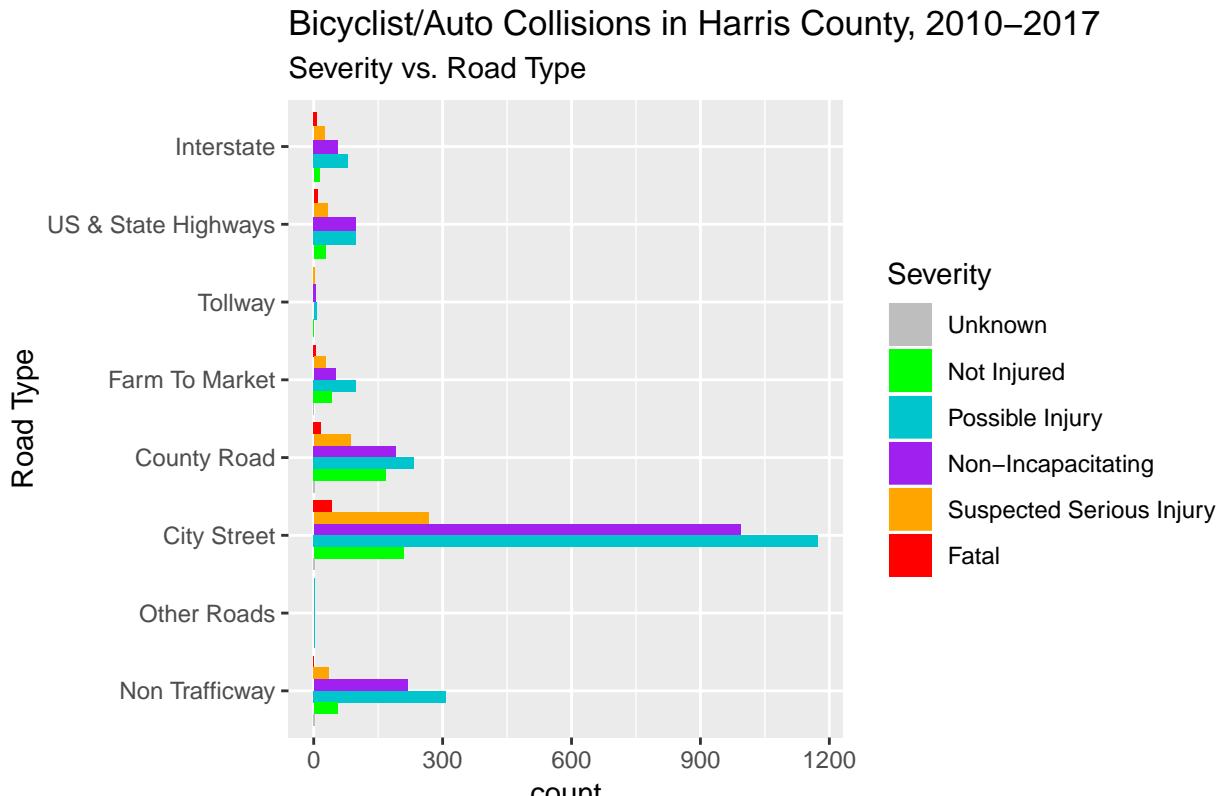
## Injuries as a function of road type

It seems obvious that if you get hit on the freeway, you are more likely to die, but let's test that hypothesis, just to be sure.

As the plot indicates, injuries on a highway are about as likely to be fatal, but most collisions occur on city streets. So, unlike pedestrians, fatality risk for bicycles does not seem to correlate to road type.

But what the heck is this “non trafficway” category of road? I think it must mostly be parking lots, with a few driveways thrown in.

```
cyclists %>% ggplot() +
  geom_bar(aes(x=RoadType, fill=Severity), position="dodge" ) +
  coord_flip() +
  scale_fill_manual("Severity",values = c("gray", "green", "turquoise3", "purple", "orange", "red")) +
  xlab("Road Type")+
  labs(title = "Bicyclist/Auto Collisions in Harris County, 2010–2017",
       subtitle = "Severity vs. Road Type",
       caption = caption) +
  theme(legend.position = "right")
```



## Data issues

Note that the data is not perfect, there are a few issues. Not all of the locations make sense, and so cannot be linked to a latitude and longitude location, limiting the analysis to a degree. Only 83.8% of the data is geocoded.

## City Streets and Bicyclists

My primary interest is looking at city streets, to see where there may be hotspots, and to try to predict which spots are most likely to see a fatality or serious injury. So for now, I will restrict analysis to “City Streets”.

Let’s consider what time of day these incidents occur. I have heard it proposed that many incidents may be due to bars letting out and people wandering into the road in an impaired state. Or perhaps the driver is impaired.

Doesn’t look like impairment is a major factor. It appears that the incidents peak during rush hour. Well, there is a minor peak at 2 AM when the bars close. And an intriguing jump at 3 PM. Is that related to school letting out? I’ll have to look at that later.

```
citycyclists <- cyclists %>% filter(RoadType=="City Street")
citycyclists$Intersection <- as.factor(citycyclists$Intersection)

citycyclists %>% mutate(HourOfDay=hour(citycyclists$Crash_DateTime)) %>%
  ggplot() +
  geom_bar(aes(x=HourOfDay), fill = "darkorchid4") +
```

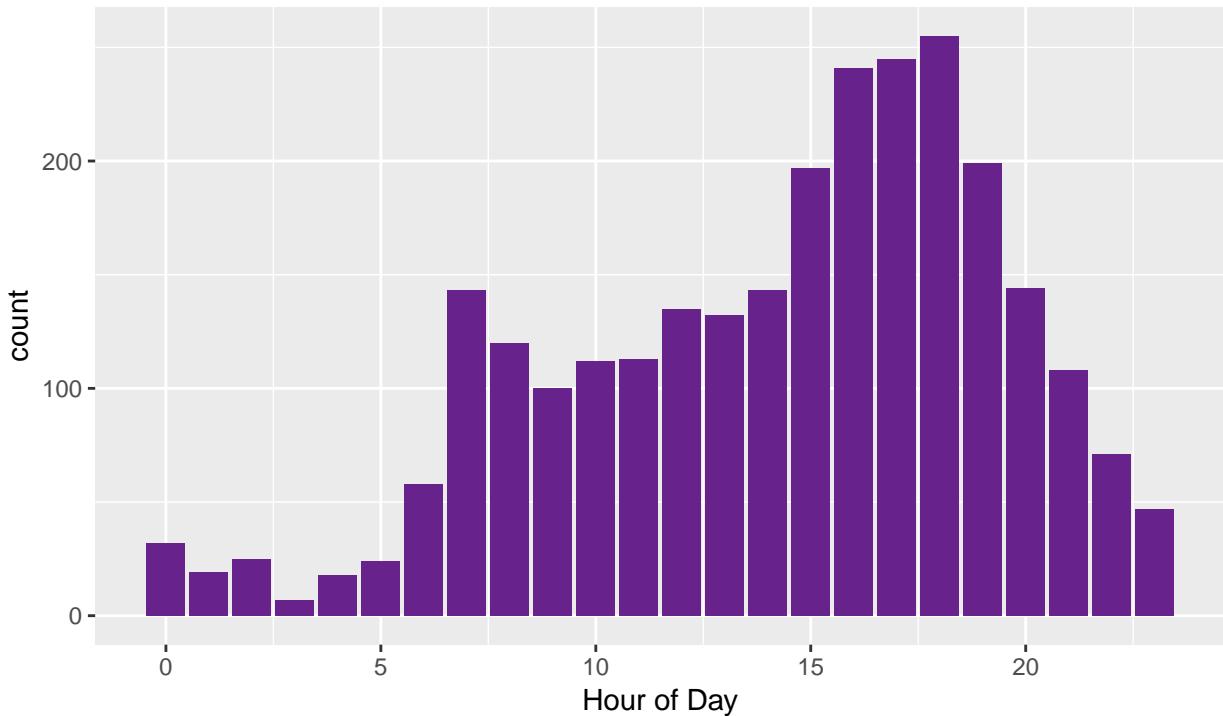
```

xlab("Hour of Day")+
labs(title = "Bicyclist/Auto Collisions, Houston City Streets, 2010-2017",
    subtitle = "Hour of Day",
    caption = caption)

```

## Bicyclist/Auto Collisions, Houston City Streets, 2010–2017

### Hour of Day



Alan Jackson, Adelie Resources, LLC 2018

### Day of Week

How about day of the week? Not surprisingly, Sunday is the low day. Other than Sunday, I would say that the other days of the week are basically equal in terms of number of incidents.

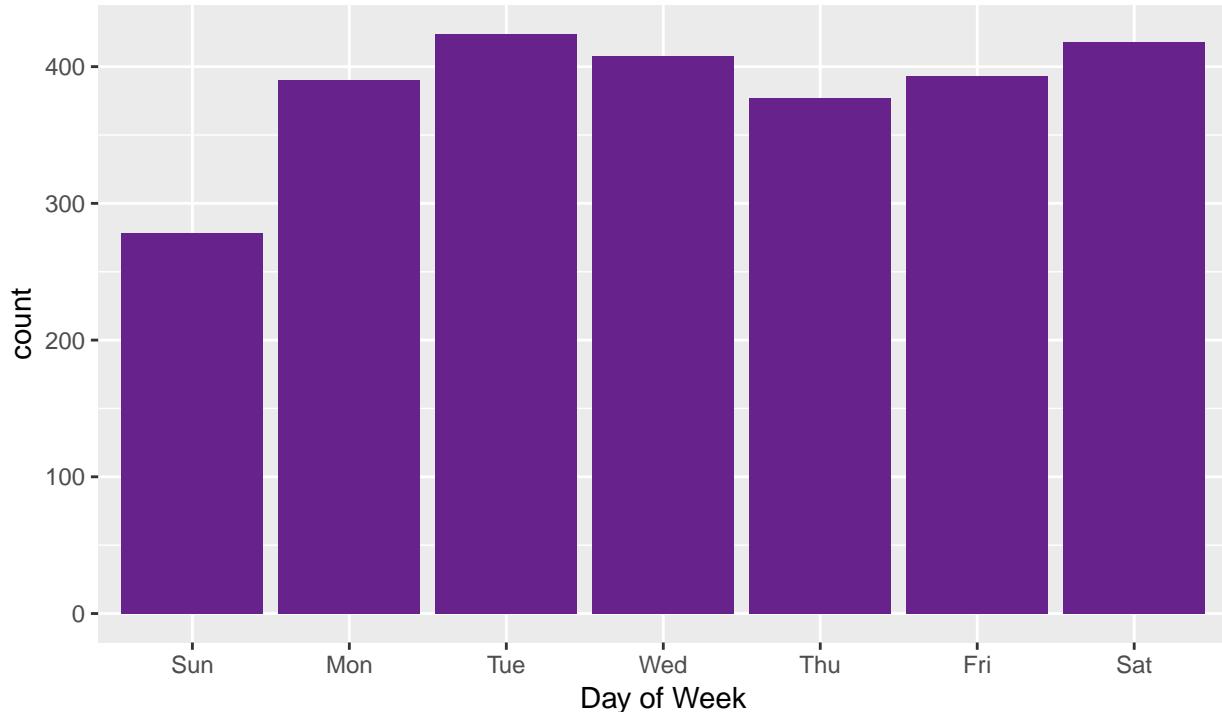
```

citycyclists %>% mutate(Daynum=wday(Crash_Date)) %>%
ggplot(aes(x=wday(Daynum, label=TRUE))) +
geom_bar(fill = "darkorchid4") +
xlab("Day of Week")+
labs(title = "Bicyclist/Auto Collisions, Houston City Streets, 2010-2017",
    subtitle = "Day of Week",
    caption = caption)

```

## Bicyclist/Auto Collisions, Houston City Streets, 2010–2017

### Day of Week



Alan Jackson, Adelie Resources, LLC 2018

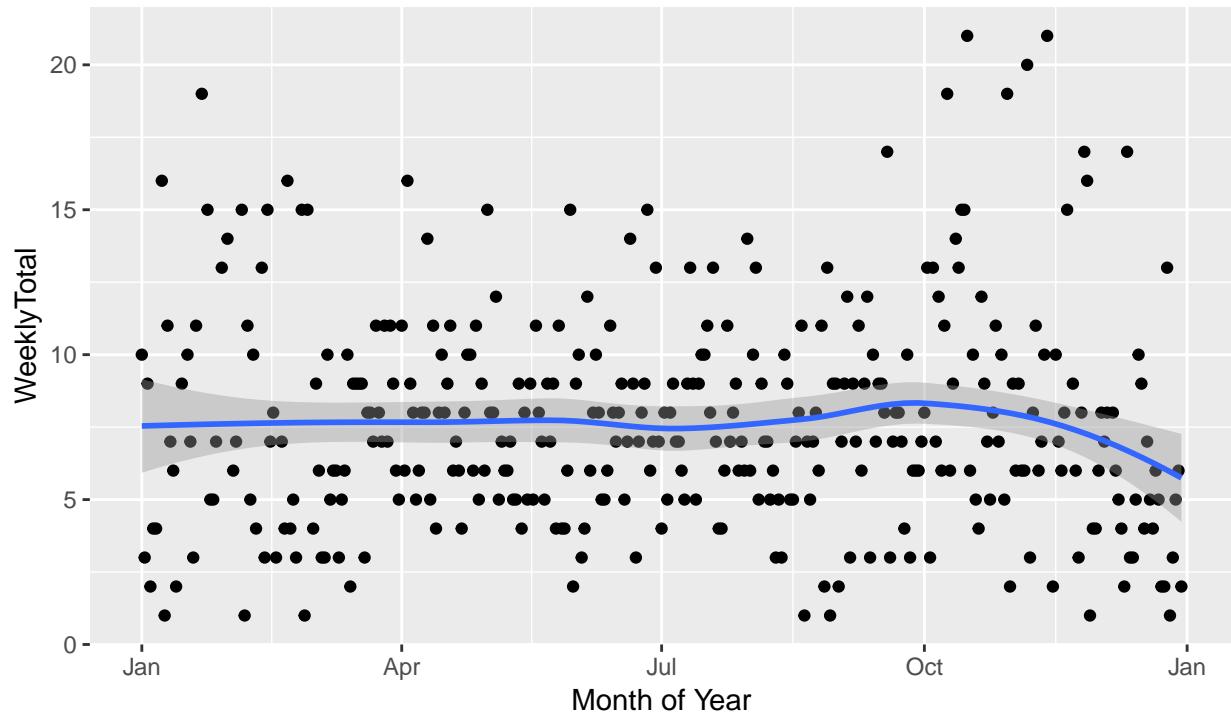
### Time of Year

Huge scatter, but really no obvious seasonal effect. Tells me that bicyclists are a dedicated bunch, and ride all year long without regard to weather.

```
citycyclists %>% mutate(WkOfYear=as.Date(paste0("2000-",month(floor_date(Crash_Date, "week")),"-",day(floor_date(Crash_Date, "week"))))) %>% group_by(WkOfYear) %>% summarise(WeeklyTotal=n()) %>% ggplot(aes(x=WkOfYear, y=WeeklyTotal)) + geom_point() + geom_smooth() + scale_x_date(date_labels = "%b") + xlab("Month of Year") + labs(title = "Bicyclist/Auto Collisions, Houston City Streets, 2010-2017", subtitle = "Month of Year", caption = caption)
```

## Bicyclist/Auto Collisions, Houston City Streets, 2010–2017

Month of Year



Alan Jackson, Adelie Resources, LLC 2018

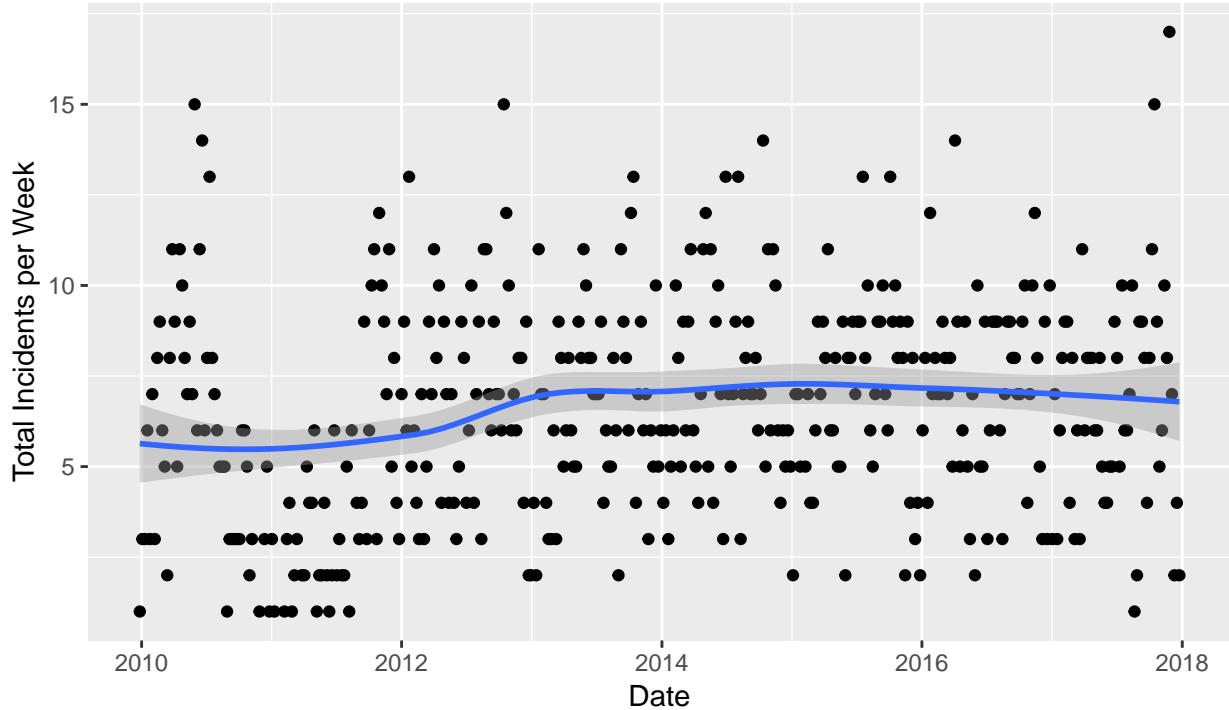
### Long-term trends - is it getting better?

I'd say that there is no obvious long-term trend. Not getting better, not getting worse. On the other hand, I suspect that the number of riders may be increasing, so a flat incident count would be good news.

```
citycyclists %>% mutate(ByWeek=floor_date(Crash_Date, "week")) %>%
  group_by(ByWeek) %>%
  summarise(WeeklyTotal=n()) %>%
  ggplot(aes(x=ByWeek, y=WeeklyTotal)) +
  geom_point() +
  geom_smooth() +
  xlab("Date") +
  ylab("Total Incidents per Week") +
  labs(title = "Bicyclist/Auto Collisions, Houston City Streets, 2010–2017",
       subtitle = "Weekly Totals",
       caption = caption)
```

## Bicyclist/Auto Collisions, Houston City Streets, 2010–2017

### Weekly Totals



Alan Jackson, Adelie Resources, LLC 2018

### Where are the worst spots for number of incidents?

The worst is a location that has racked up 7 incidents in 7 years - about one incident per year. 108 locations have seen two incidents, 20 locations have seen three. If one were looking for a way to prioritize work, this might be a good place to start.

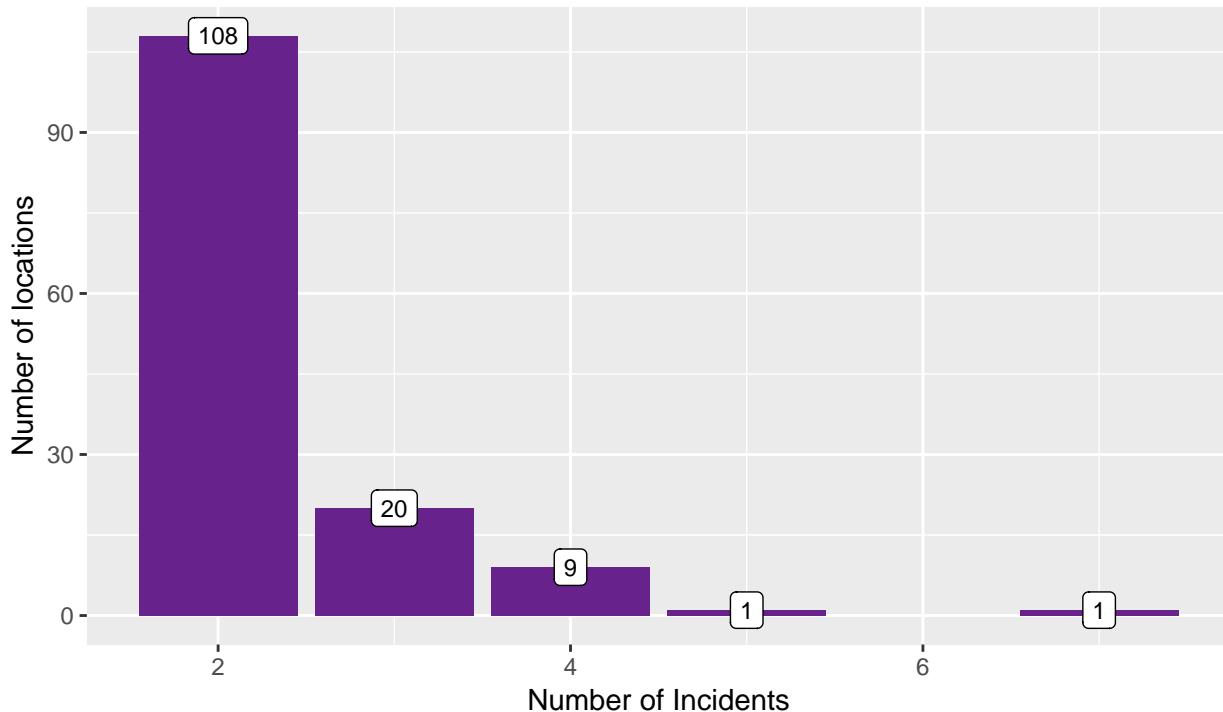
```
citycyclistslatlng <- citycyclists %>%
  filter(!is.na(Latitude))

badspots <- citycyclistslatlng %>%
  group_by(latlon) %>%
  summarise(total=n()) %>%
  filter(total>1)

badspots %>%
  ggplot(aes(x=total)) +
  geom_bar(fill = "darkorchid4") +
  geom_label(aes(label=..count..), stat='count', size = 3) +
  labs(title = "Bicyclist/Auto Collisions in Harris County, 2010–2017",
       subtitle = "Number of locations with more than one incident",
       caption = caption) +
  xlab("Number of Incidents")+
  ylab("Number of locations")
```

## Bicyclist/Auto Collisions in Harris County, 2010–2017

Number of locations with more than one incident



Alan Jackson, Adelie Resources, LLC 2018

## What controls the Severity?

What controls how serious an incident turns out to be? Most probably the speed of the vehicle. So let's take a look at that.

Serious injuries clearly climb with speed. Not injured has a curious spike at high speeds, but the likely reporting bias for those numbers is so large I don't think any conclusions can be made. And fatality numbers are small enough that any conclusions from those would be suspect.

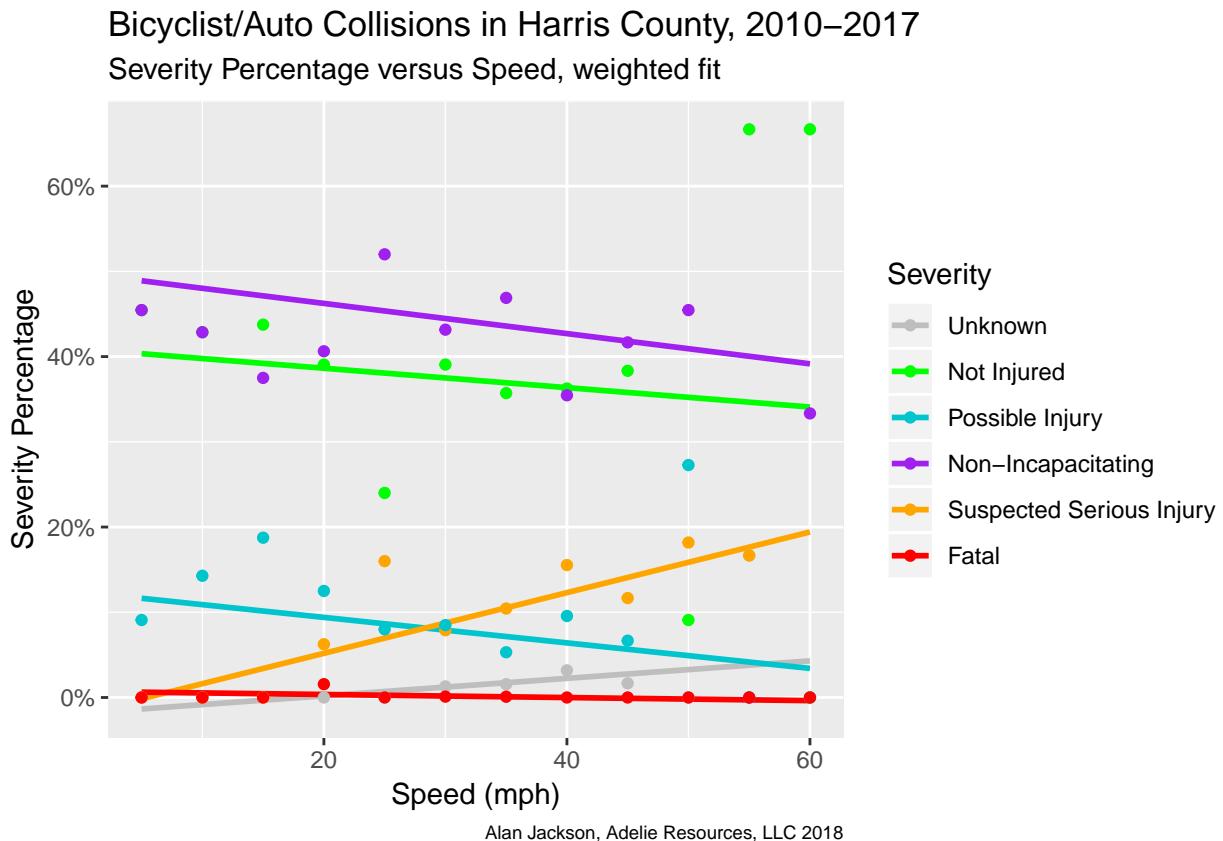
```
# round speed to nearest 5 mph

citycyclists %>% mutate(speed=as.integer((Crash_Speed_Limit+2.5)/5)*5) %>%
  filter(speed>0) %>%
  group_by(speed) %>%
  summarise(totbyspeed=n(),
            fatal=sum(Crash_Sev_ID==4),
            serious=sum(Crash_Sev_ID==1),
            injury=sum(Crash_Sev_ID==2),
            maybe=sum(Crash_Sev_ID==3),
            okay=sum(Crash_Sev_ID==5),
            unk=sum(Crash_Sev_ID==0)) %>%
  mutate(fatalpct=fatal/totbyspeed,
        seriouspct=serious/totbyspeed,
        injurypct=injury/totbyspeed,
        maybepct=maybe/totbyspeed,
        okaypct=okay/totbyspeed,
```

```

    unkpct=unk/totbyspeed
) %>%
select(speed, totbyspeed, fatalpct:unkpct) %>%
rename(Fatal=fatalpct, Serious_Injury=susceptiblepct, Non_Incapacitating=injurypct, Possible_Injury=maybeinj)
gather(Severity, Pct, Fatal:Unknown) %>%
ggplot() +
geom_smooth(aes(x=speed, y=Pct, color=Severity, weight=totbyspeed), se=FALSE, method='lm')+
geom_point(aes(x=speed, y=Pct, color=Severity))+
scale_color_manual("Severity", values = c("gray", "green", "turquoise3", "purple", "orange", "red"), labels=c("Unknown", "Not Injured", "Possible Injury", "Non-Incapacitating", "Suspected Serious Injury", "Fatal"))
labs(title = "Bicyclist/Auto Collisions in Harris County, 2010–2017",
subtitle = "Severity Percentage versus Speed, weighted fit",
caption = caption) +
xlab("Speed (mph)")+
ylab("Severity Percentage") +
scale_y_continuous(labels=scales::percent) +
theme(legend.position = "right")

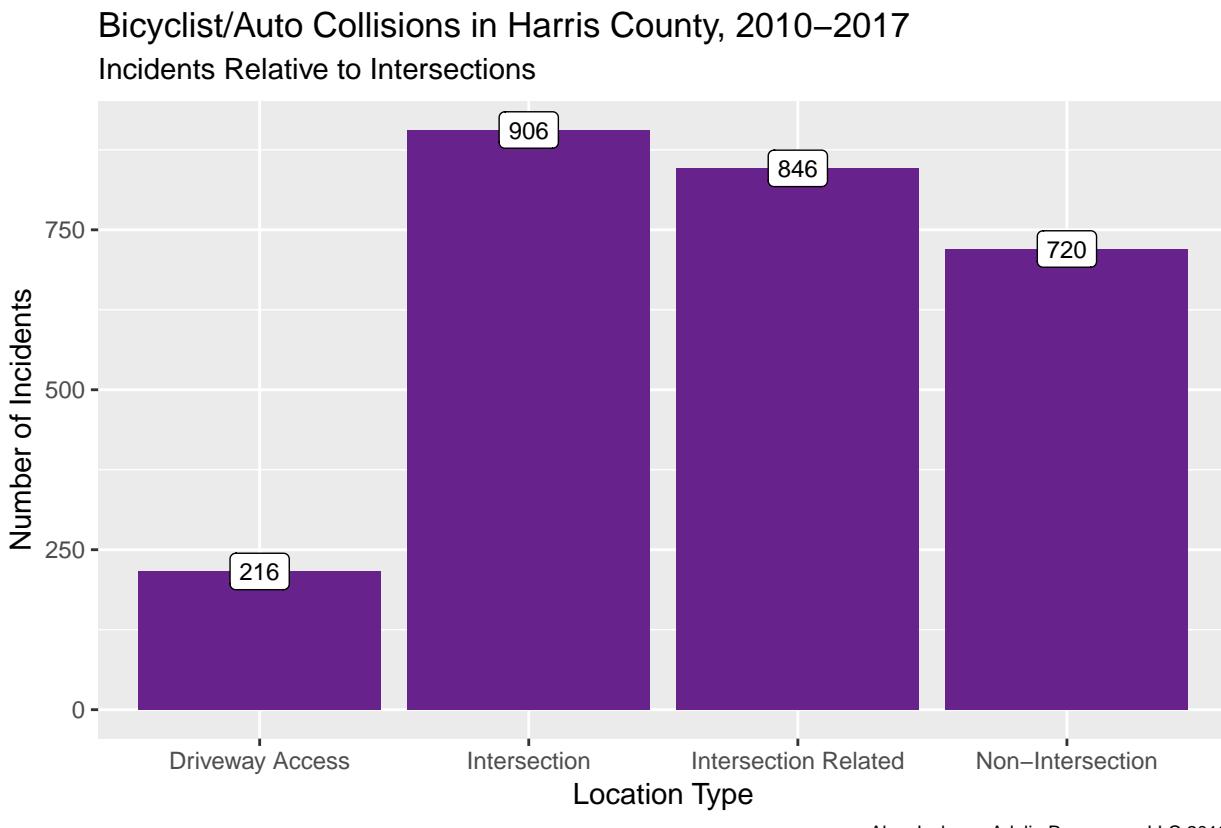
```



## What about Intersections?

Is there a pattern related to where the incident took place? At an intersection, or somewhere else? Are people hit crossing an intersection, or in the middle of the block? It appears that about 2/3 of incidents are associated with intersections. My bet is that a big fraction are cars turning. Which may imply that bicycle lanes, while something I personally like, may have limited impact on safety.

```
citycyclists %>%
  ggplot(aes(x=Intersection)) +
  geom_bar(fill = "darkorchid4") +
  geom_label(aes(label=..count..), stat='count', size = 3) +
  labs(title = "Bicyclist/Auto Collisions in Harris County, 2010-2017",
       subtitle = "Incidents Relative to Intersections",
       caption = caption) +
  xlab("Location Type")+
  ylab("Number of Incidents")
```



## How does the type of crossing relate to the injury severity?

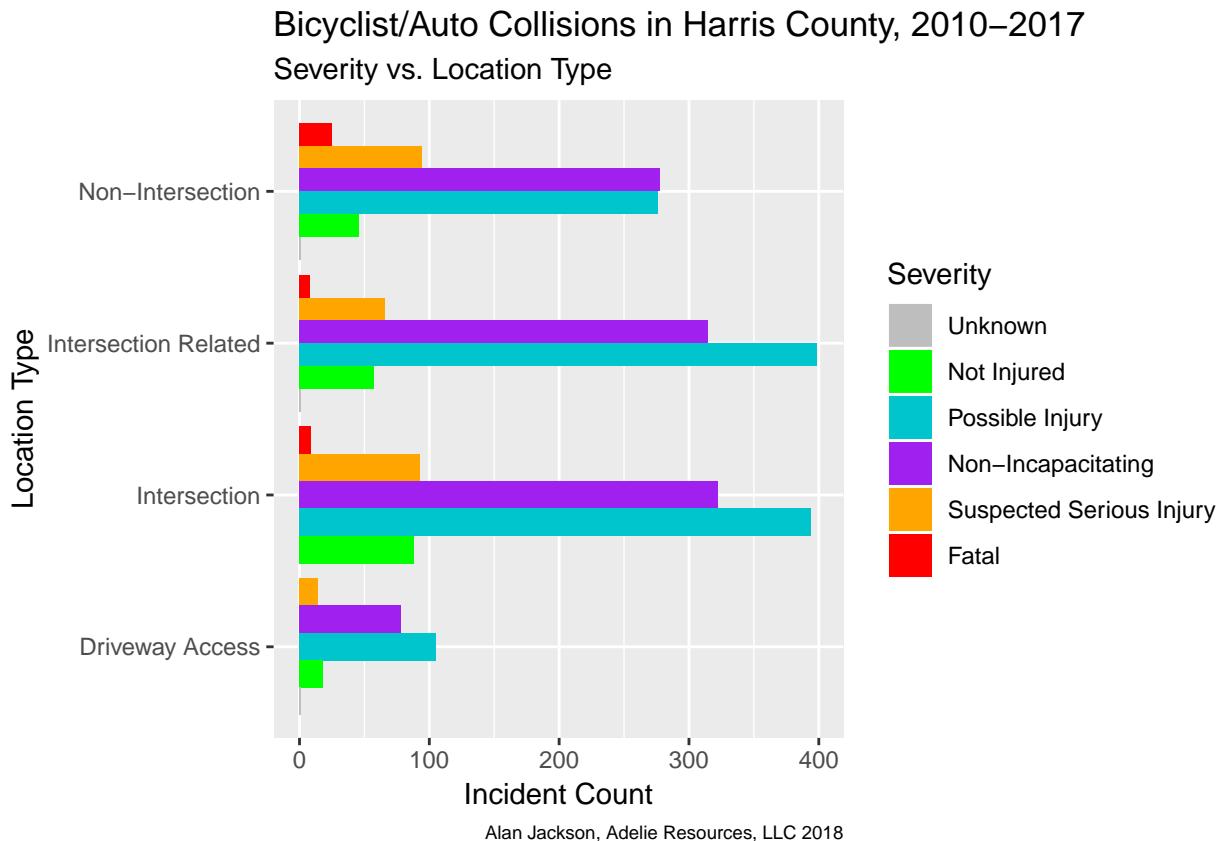
The severity of incidents away from intersections is clearly worse than those at or near intersections. So maybe bike lanes are a good idea after all. While the number of incidents may be less, cars are likely traveling faster, and speed is bad.

```
citycyclists %>% filter(Intersection!="Not Reported") %>%
  ggplot() +
  geom_bar(aes(x=Intersection, fill=Severity), position="dodge" ) +
  coord_flip() +
  scale_fill_manual("Severity",values = c("gray", "green", "turquoise3", "purple", "orange", "red")) +
  xlab("Location Type")+
  ylab("Incident Count")+
  labs(title = "Bicyclist/Auto Collisions in Harris County, 2010-2017",
       subtitle = "Severity vs. Location Type",
```

```

    caption = caption) +
  theme(legend.position = "right")

```



## Another look at location type and injury severity

Here we see quite clearly - after consolidating “Intersection Related” and “Intersection” together - that not being at an intersection is much more dangerous. The likelihood of death is about 20% greater at non-intersections. Serious injuries are more likely to occur at intersections however - probably related to speed.

```

citycyclists %>%
  filter(Severity!="Unknown") %>%
  group_by(Severity, Intersection) %>%
  summarise(total=n()) %>%
  mutate(Intersection=gsub(" ", "_", Intersection, fixed=TRUE)) %>%
  mutate(Intersection=gsub("-", "_", Intersection, fixed=TRUE)) %>%
  spread(Intersection, total) %>%
  mutate(mysum=(Intersection + Intersection_Related + Non_Intersection)) %>%
  mutate(Intersection=Intersection + Intersection_Related,
        Intersection=Intersection/mysum, Non_Intersection=Non_Intersection/mysum) %>%
  select(Severity, Intersection, Non_Intersection) %>%
  gather(Location, Probability, -Severity) %>%
  ggplot(aes(x=Location, y=Probability)) +
  geom_bar(aes(fill=Severity), position="dodge", stat='identity') +
  scale_fill_manual("Severity", values = c("green", "turquoise3", "purple", "orange", "red"))

```

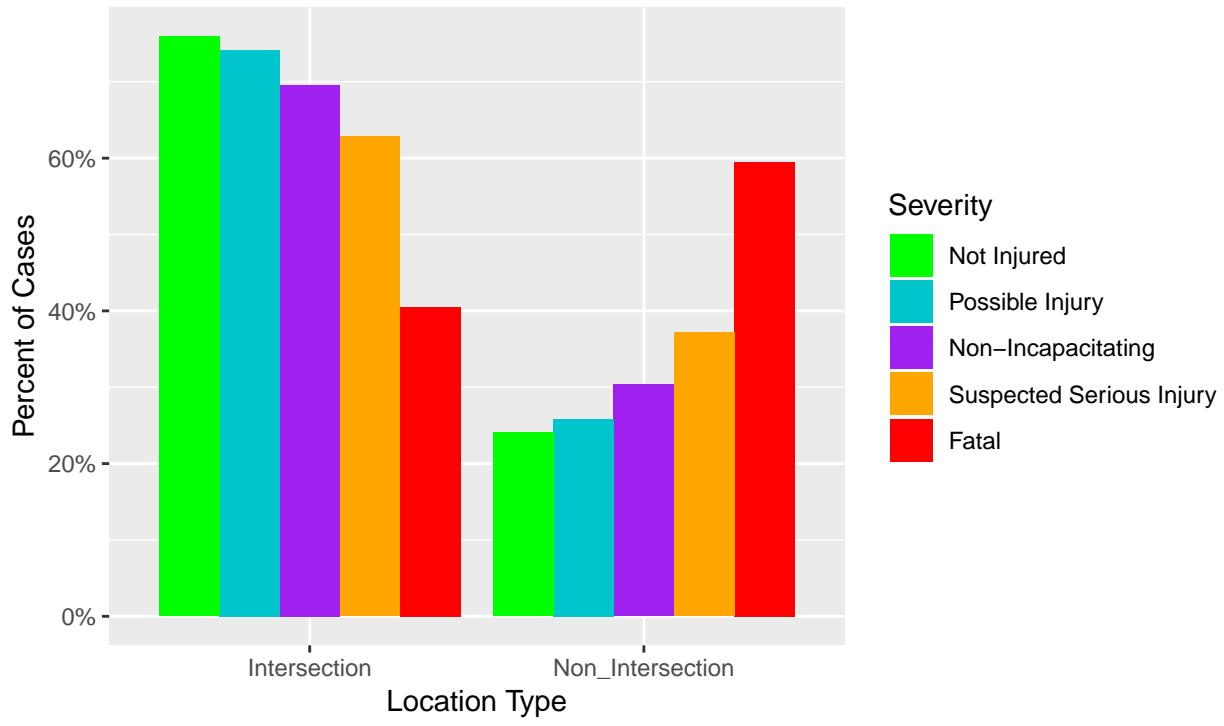
```

xlab("Location Type")+
ylab("Percent of Cases")+
scale_y_continuous(labels=scales::percent) +
labs(title = "Bicyclist/Auto Collisions in Harris County, 2010–2017",
    subtitle = "Severity vs. Location Type",
    caption = caption) +
theme(legend.position = "right")

```

## Bicyclist/Auto Collisions in Harris County, 2010–2017

### Severity vs. Location Type



Alan Jackson, Adelie Resources, LLC 2018

## Where are most Incidents?

Lake Houston area looks pretty dangerous.

```

# Locations with 5 or more incidents
mytable <- citycyclistslatlng %>%
  group_by(latlon, Rpt_Street_Name) %>%
  summarise(total=n()) %>%
  filter(total>2) %>%
  arrange(desc(total)) %>%
  ungroup() %>%
  select(Street=Rpt_Street_Name, Total=total)
print(xtable(mytable, auto=TRUE), type="html")

```

Street

Total

1

MAGNOLIA COVE

4

2

LAKE HOUSTON W

4

3

CHIMNEY ROCK

4

4

FAIRMONT

3

5

LAKE HOUSTON W

3

6

SANDY FORKS

3

7

BROOK SHADOW

3

8

NORTHPARK

3

9

SPENCER

3

10

EDGEBROOK

3

11

MACGREGOR

3

12

MAIN

3

```

13
ELLA
3
14
GESSNER
3
15
HONEYWELL
3
16
BELLAIRE
3
17
BELLAIRE
3
18
DAIRY ASHFORD
3

```

## Maps

Let's make some maps.

If we plot all locations that had three or more incidents, we produce the map below. A Large concentration of incidents downtown - that's where many bicyclists are - but also some of the largest values are outside of downtown.

```

zoom <- 10
center <- c(-95.4, 29.8)
gmap = get_map(location=center, source="google", zoom=zoom)

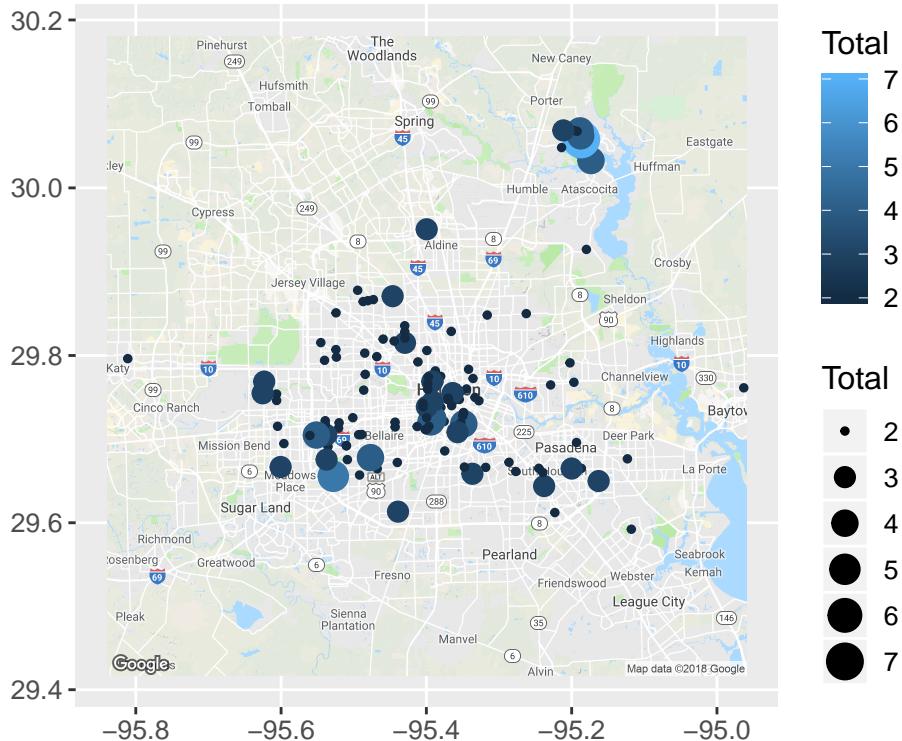
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.8,-95.4&zoom=10&size=640x640&
bad_bg <-
  citycyclistslatlng %>%
  group_by(latlon, Latitude, Longitude) %>%
  summarise(total_bg=n()) %>%
  filter(total_bg>1)

ggmap(gmap, extent='normal', maprange=FALSE, show.legend=FALSE) %+% bad_bg + aes(x = Longitude, y = Lat
  geom_point(data=bad_bg, aes(size=total_bg,color=total_bg)) +
  scale_color_continuous(guide="colorbar", name="Total") +
  scale_alpha(range = c(0.00, 0.35), guide = FALSE) +
  labs(title="Collisions 2010-2017", subtitle="Two or more incidents only", size="Total") +
  theme(legend.position = "right", axis.title = element_blank(), text = element_text(size = 12))

```

## Collisions 2010–2017

### Two or more incidents only



### Serious Incidents only

Let's look at locations that had a fatality or serious injury.

Few obvious single locations pop out, however, a number streets resemble a string of pearls, indicating that there are stretches of some streets that seem to be particularly dangerous.

```

zoom <- 11
center <- c(-95.4, 29.8)
gmap = get_map(location=center, source="google", zoom=zoom)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.8,-95.4&zoom=11&size=640x640&
badspots <-
  citycyclistslatlng %>%
  mutate(Injury = case_when(
    Severity %in% Severity ~ "Serious",
    TRUE ~ "Non-Serious"
  )) %>%
  filter(Injury=="Serious") %>%
  filter(
    between(Latitude, HoustonBound[2], HoustonBound[4]) &
    between(Longitude,HoustonBound[1], HoustonBound[3]) ) %>%
  group_by(latlon, Latitude, Longitude) %>%
  summarise(Total=n())

ggmap(gmap, extent='normal', maprange=FALSE, show.legend=FALSE) %+% badspots + aes(x = Longitude, y = L
  
```

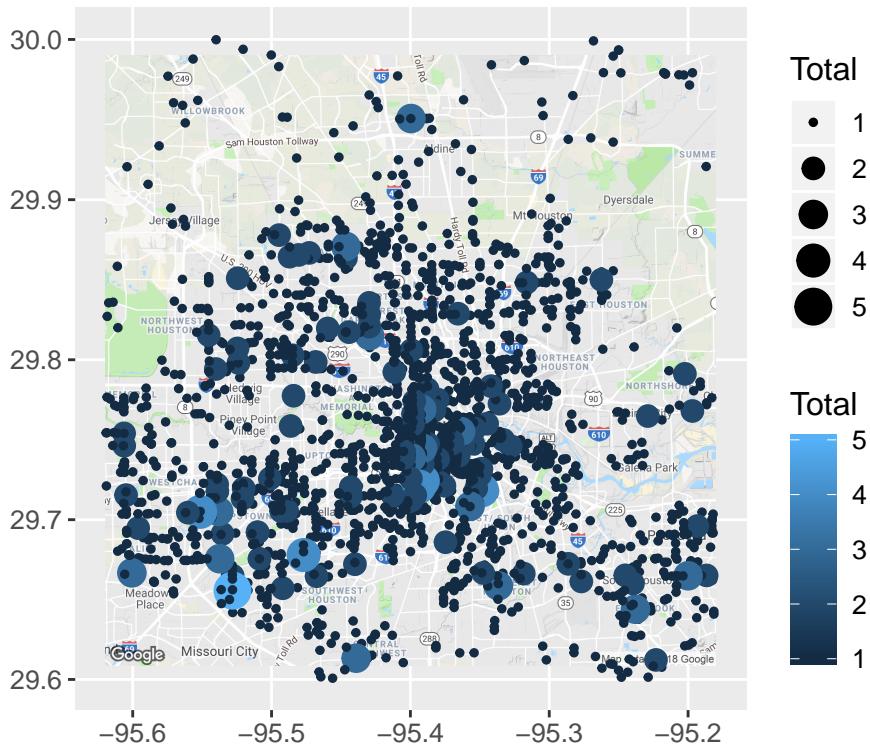
```

geom_point(data=badspots, aes(size=Total,color=Total)) +
  scale_color_continuous(guide="colorbar", name="Total") +
  scale_alpha(range = c(0.00, 0.35), guide = FALSE) +
  labs(title="Collisions 2010-2017", subtitle="Fatal or Serious Only", size="Total") +
  theme(legend.position = "right", axis.title = element_blank(), text = element_text(size = 12))

```

## Collisions 2010–2017

### Fatal or Serious Only



```

Serious <- c("Suspected Serious Injury", "Fatal")

badspots <-
  citycyclistslatlng %>%
  mutate(Injury = case_when(
    Severity %in% Serious ~ "Serious",
    TRUE ~ "Non-Serious"
  )) %>%
  mutate(TimeOfDay = case_when(
    between(hour(Crash_DateTime), 7, 20) ~ "Day",
    TRUE ~ "Night"
  )) %>%
  group_by(latlon, Latitude, Longitude, Injury, TimeOfDay) %>%
  summarise(total=n())

facetmap <- function(){
  badspots_bg <- bad_spots %>%
    filter(
      between(Latitude, Center[2]-0.01, Center[2]+0.01) &
      between(Longitude, Center[1]-0.01, Center[1]+0.01))

```

```

badmap <- badspots %>% filter(
  between(Latitude,Center[2]-0.01, Center[2]+0.01) &
  between(Longitude,Center[1]-0.01, Center[1]+0.01))

ggmap(gmap, extent='normal', maprange=FALSE, show.legend=FALSE) %+% badmap + aes(x = Longitude, y = Latitude)
  geom_point(data=badspots_bg, aes(size=total_bg, x = Longitude, y = Latitude), fill="yellow", shape=21) +
  geom_point(data=badmap, aes(size=total, x = Longitude, y = Latitude, color=total)) +
  facet_wrap(TimeOfDay ~ Injury) +
  guides(fill = FALSE) + # to remove the legend
  theme_bw() + # for clean look overall
  scale_color_continuous(guide="colorbar", name="Subtotal") +
  scale_alpha(range = c(0.00, 0.35), guide = FALSE) +
  labs(title=paste("Bicyclist Collisions", Place,"2010-2017"), size="Total", subtitle="Yellow for total")
  theme(legend.position = "right", axis.title = element_blank(), text = element_text(size = 12))
}

}

```

## Downtown

Not surprisingly, almost all incidents downtown occur during the day. St. Joseph Parkway, especially at San Jacinto, seems to be a bad spot. Smith around Tranquility Park, and a block away, Rusk at Louisiana is a bad spot.

```

Center <- c(-95.367, 29.757)
Place <- "Downtown"

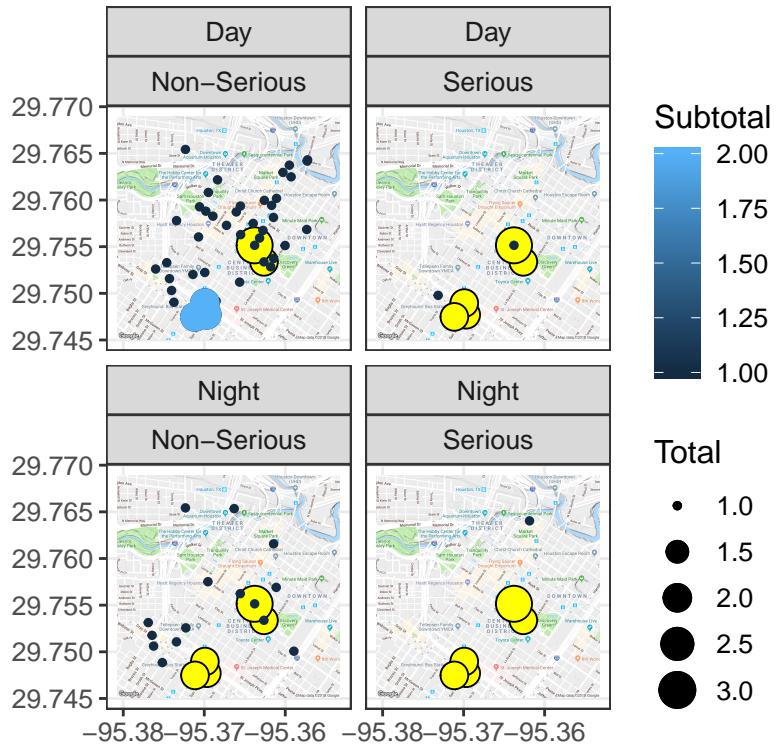
gmap = get_map(location=Center, source="google", zoom=15)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.757,-95.367&zoom=15&size=640x480
facetmap()

```

## Bicyclist Collisions Downtown 2010–2017

Yellow for total incidents of 3 or more, blue for subset by time and



### Montrose/Midtown

Montrose is bad, day and night. It is also bad for pedestrians. Confusing intersections, heavy traffic, plenty of bars and restaurants - all add up to injuries and death.

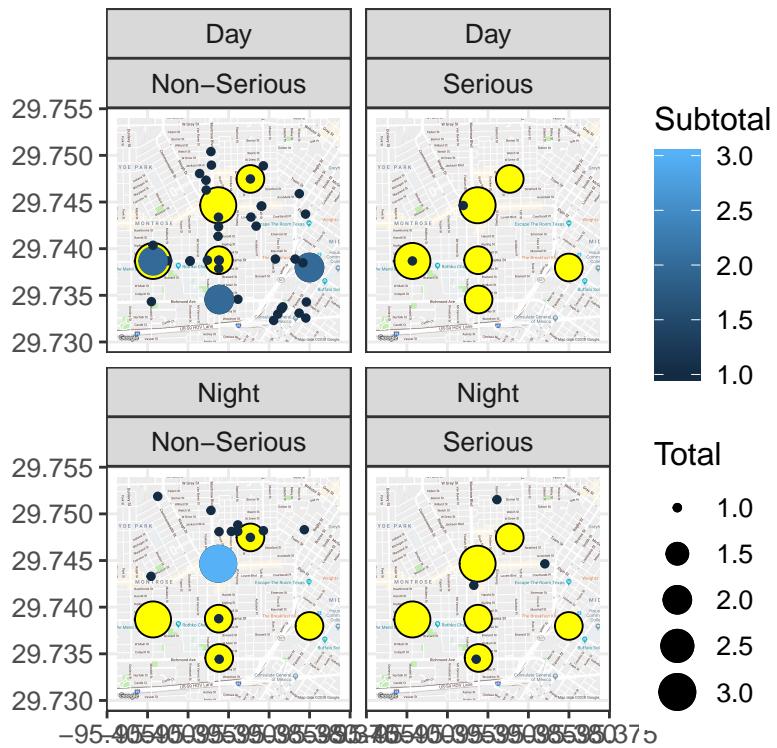
```
Center <- c(-95.39, 29.742)
Place <- "Midtown"

gmap = get_map(location=Center, source="google", zoom=15)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.742,-95.39&zoom=15&size=640x640
facetmap()
```

## Bicyclist Collisions Midtown 2010–2017

Yellow for total incidents of 3 or more, blue for subset by time and



## University Place

The poster child for bicycle fatalities due to the train collisions, but other serious incidents have also occurred.

```
Center <- c(-95.405, 29.717)
```

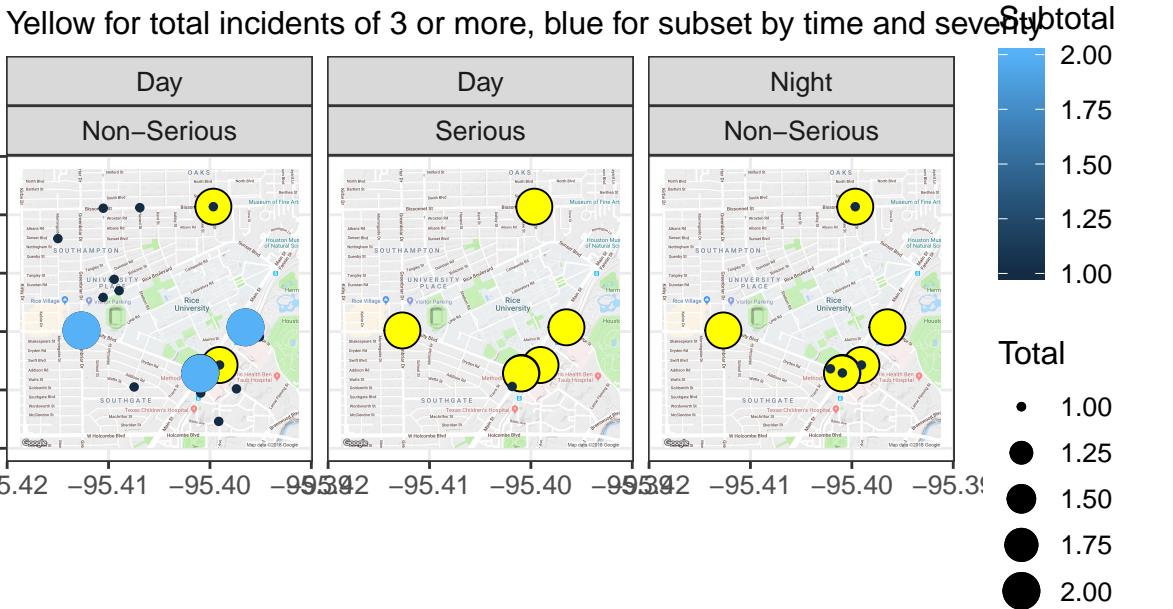
```
Place <- "University Place"
```

```
gmap = get_map(location=Center, source="google", zoom=15)
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.717,-95.405&zoom=15&size=640x640
```

```
facetmap()
```

## Bicyclist Collisions University Place 2010–2017



## Chinatown

```

Center <- c(-95.545, 29.710)
Place <- "Sharpstown"

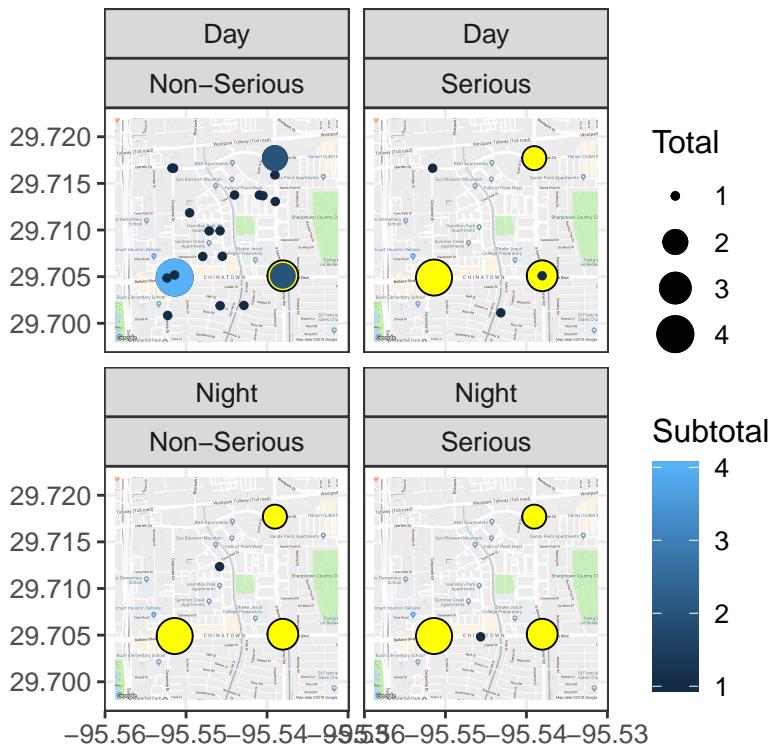
gmap = get_map(location=Center, source="google", zoom=15)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.71,-95.545&zoom=15&size=640x640
facetmap()

```

## Bicyclist Collisions Sharpstown 2010–2017

Yellow for total incidents of 3 or more, blue for subset by time and



### Atascosita

Will Clayton Parkway appears to be fairly dangerous for bicycles, with four incidents, one serious.

```
Center<- c(-95.208, 29.977)
```

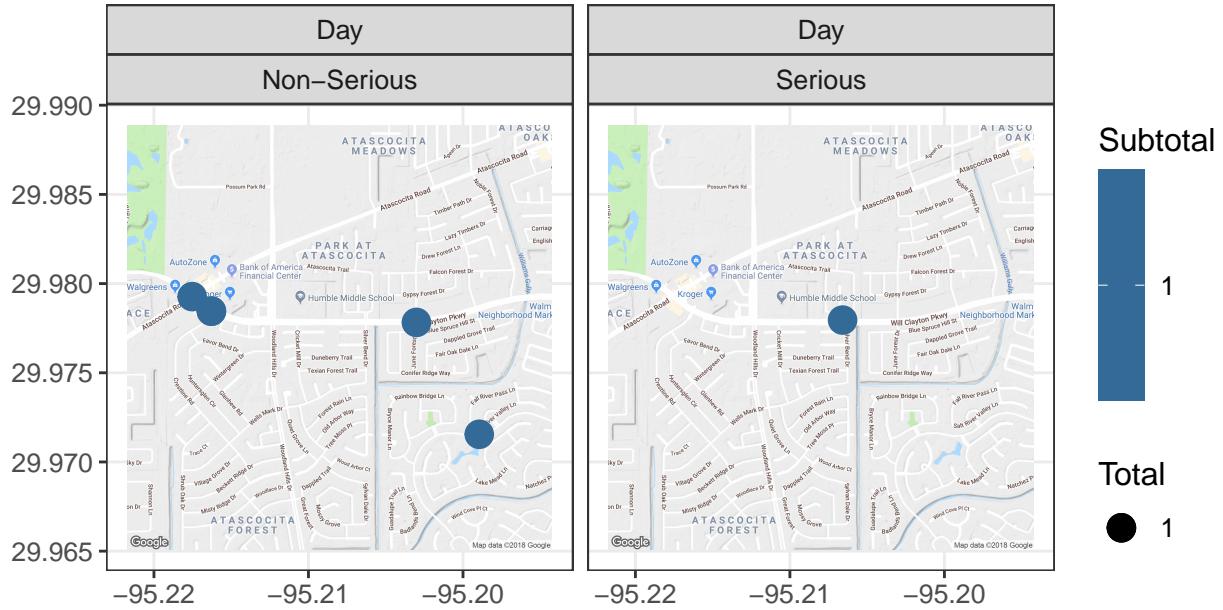
```
Place <- "Atascosita"
```

```
gmap = get_map(location=Center, source="google", zoom=15)
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.977,-95.208&zoom=15&size=640x640
```

## Bicyclist Collisions Atascocita 2010–2017

Yellow for total incidents of 3 or more, blue for subset by time and severity



## Airline

Airline and some other streets in the area form strings of pearls.

```
Center<- c(-95.384, 29.833)
```

```
Place <- "Airline"
```

```
gmap = get_map(location=Center, source="google", zoom=15)
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=29.833,-95.384&zoom=15&size=640x640
```

```
facetmap()
```

## Bicyclist Collisions Airline 2010–2017

Yellow for total incidents of 3 or more, blue for subset by time and severity

