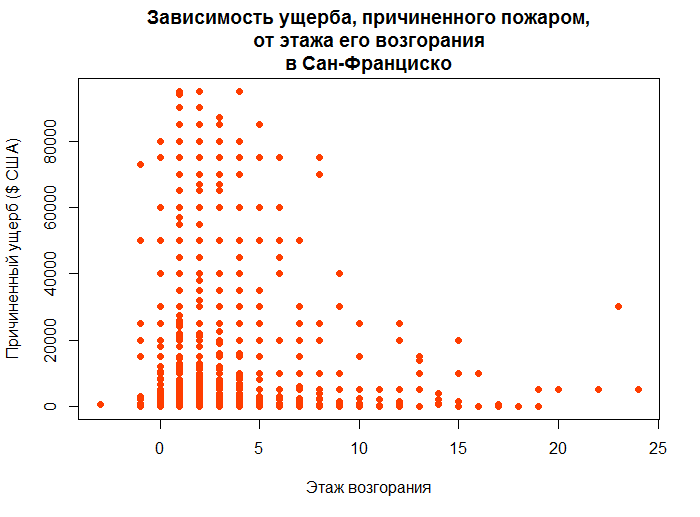
**Задание 3**

*Чтение данных из двумерного массива и построение облака точек*

1. Выполните следующие действия:
2. Создайте текстовый файл с двумерным массивом данных (данные должны иметь экономический, физический или иной смысл).
3. Загрузите данные в пакет R.
4. Постройте облако точек для загруженных данных.
5. Сделайте выводы (на основе визуальной оценки) о характере зависимости признаков.
6. Оформите результаты исследования в виде Word-файла и прикрепите его здесь.

**Был выбран достаточно большой data frame, содержащий данные о пожарах в Сан-Франциско.**

Получившееся облако точек:



**Вывод**

К сожалению, не удалось увидеть какой-либо однозначной линейной зависимости в выбранных данных. Более того, для ее поиска необходимо корректнее определять выборку данных: например, выбрать пожары, произошедшие только в пятиэтажных домах.

Также в скрипте программы приведены мои различные попытки найти возможные зависимости (иногда мне казалось, что зависимость должна явно проявиться) между различными величинами, приведенными в этом data frame’е. Однако, мне их так и не удалось найти.

P.S. используемые мной данные, а также множество других data frame’ов можно найти по [ссылке](https://data.sfgov.org/).

Скрипт программы на R:

setwd("C:\\Users\\user\\Desktop\\rwd")

# используем эту опцию, чтобы видеть числа на графике в удобном формате

# т.е. НЕ в экспоненциальной записи

options(*scipen*=*5*)

*fire\_data* **<-** read.csv("Fire\_incidents.csv")

*time.format* **<-** "%m/%d/%Y %I:%M:%S %p"

# попробуем узнать, есть ли линейная зависимость между

# величиной ущерба от пожара и днем, когда случился пожар

*fire.day.to.loss* **<-** *fire\_data*[c(*1*,*5*)]

length(*fire.loss*[,*1*])

*fire.day.to.loss*[,*1*] **<-** as.factor(weekdays(as.Date(*fire.day.to.loss*[,*1*], *format*="%m/%d/%Y"), *abbreviate*=*TRUE*))

levels(*fire.day.to.loss*[,*1*]) **<-** list(*Пн*=*1*, *Вт*=*2*, *Ср*=*3*, *Чт*=*4*, *Пт*=*5*, *Сб*=*6*, *Вс*=*7*)

*fire.day.to.loss*[,*1*] **<-** as.numeric(*fire.day.to.loss*[,*1*])

plot(*fire.day.to.loss*[,*1*], *fire.day.to.loss*[,*2*], *col*="blue", *type*="p", *pch*=*16*, *main*="Зависимость ущерба, причиненного пожаром, от дня недели")

# есть ли линейная зависимость между продолжительностью тушения пожара

# и величиной ущерба от него

*fire.time.to.loss* **<-** *fire\_data*[*5*]

*fire.time.to.loss*[,*2*] **<-** mapply(**function**(*start*, *end*) difftime(strptime(*end*, *format*=*time.format*), strptime(*start*, *format*=*time.format*), *units*="mins"), *fire\_data*[,*3*], *fire\_data*[,*4*])

colnames(*fire.time.to.loss*) **<-** c("Loss", "Time")

plot(*fire.time.to.loss*[, *2*], *fire.time.to.loss*[, *1*],

*col*="#FF3D00", *type*="p",

*pch*=*16*,

*main*="Зависимость ущерба, причиненного пожаром, от продолжительности его тушения в Сан-Франциско",

*xlab*="Продолжительность тушения пожара (мин)",

*ylab*="Причиненный ущерб ($ США)")

*fire\_floor\_data* **<-** read.csv("fire.floors.csv")

# есть ли линейная зависимость между этажом, на котором начался пожар

# и продолжительностью его тушения

*fire.floor.to.time* **<-** *fire\_floor\_data*[*7*]

*fire.floor.to.time*[,*2*] **<-** mapply(**function**(*start*, *end*) difftime(strptime(*end*, *format*=*time.format*), strptime(*start*, *format*=*time.format*), *units*="mins"), *fire\_floor\_data*[,*3*], *fire\_floor\_data*[,*4*])

plot(*fire.floor.to.time*[,*1*], *fire.floor.to.time*[,*2*],

*col*="#FF3D00", *type*="p",

*pch*=*16*,

*main*="Зависимость продолжительности тушения пожара от этажа его возгорания в Сан-Франциско",

*xlab*="Этаж возгорания",

*ylab*="Продолжительность тушения пожара (мин)")

# есть ли линейная зависимость между временем реагирования пожарных

# и продолжительностью его тушения

*fire.reaction.to.time* **<-** data.frame()

*fire.reaction.to.time* **<-** mapply(**function**(*start*, *end*) difftime(strptime(*end*, *format*=*time.format*), strptime(*start*, *format*=*time.format*), *units*="mins"), *fire\_floor\_data*[,*2*], *fire\_floor\_data*[,*3*])

*fire.reaction.to.time* **<-** data.frame(*fire.reaction.to.time*)

*fire.reaction.to.time*[,*2*] **<-** mapply(**function**(*start*, *end*) difftime(strptime(*end*, *format*=*time.format*), strptime(*start*, *format*=*time.format*), *units*="mins"), *fire\_floor\_data*[,*3*], *fire\_floor\_data*[,*4*])

*fire.reaction.to.time* **<-** na.omit(*fire.reaction.to.time*)

summary(*fire.reaction.to.time*)

colnames(*fire.reaction.to.time*) **<-** c("Reaction", "Time")

plot(*fire.reaction.to.time*[,*1*], *fire.reaction.to.time*[,*2*],

*col*="#FF3D00", *type*="p",

*pch*=*16*,

*main*="Зависимость продолжительности тушения пожара от времени реагирования пожарных в Сан-Франциско",

*xlab*="Время реагирования (мин)",

*ylab*="Продолжительность тушения пожара (мин)")

# есть ли линейная зависимость между этажом, на котором начался пожар

# и продолжительностью его тушения

remove(*fire.floor.to.loss*)

*fire.floor.to.loss* **<-** *fire\_floor\_data*[*7*]

*fire.floor.to.loss*[,*2*] **<-** *fire\_floor\_data*[*6*]

*fire.floor.to.loss*

*fire.floor.to.loss* **<-** na.omit(*fire.floor.to.loss*)

na.fail(*fire.floor.to.loss*)

colnames(*fire.floor.to.loss*) **<-** c("Floor", "Loss")

*fire.floor.to.loss* **<-** subset(*fire.floor.to.loss*, *Floor* **<** *25*, *Loss* **<** *100000*)

plot(*fire.floor.to.loss*[,*1*], *fire.floor.to.loss*[,*2*],

*col*="#FF3D00", *type*="p",

*pch*=*16*,

*main*="Зависимость ущерба, причиненного пожаром,\nот этажа его возгорания\nв Сан-Франциско",

*xlab*="Этаж возгорания",

*ylab*="Причиненный ущерб ($ США)")