CIND820 Capstone

Marlene Martins

March 29, 2021

install.packages(‘FNN’) install.packages(‘RCurl’) install.packages(‘MASS’) install.packages(‘leaps’) library(MASS) # stepwise regression library(leaps) # all subsets regression

Dataset: <http://vincentarelbundock.github.io/> Rdatasets/csv/Ecdat/Computers.csv

library(MASS) # stepwise regression   
library(leaps) # all subsets regression  
library(RCurl)  
  
  
#c\_prices <- read.csv(file="http://vincentarelbundock.github.io/Rdatasets/csv/Ecdat/Computers.csv")  
covid\_dat <- read.csv(file="Q:/BI Team/Credit Card Reporting/Projects/ADH0820/Data/COVID\_NBH\_Summaryo.csv")  
  
head(covid\_dat)

## NEIGHBOURHOOD\_ID neighbourhood\_name TOT\_HOSPITALIZED  
## 1 1 West Humber-Clairville 128  
## 2 2 Mount Olive-Silverstone-Jamestown 127  
## 3 3 Thistletown-Beaumond Heights 54  
## 4 4 Rexdale-Kipling 27  
## 5 5 Elms-Old Rexdale 24  
## 6 6 Kingsview Village-The Westway 71  
## TOT\_IN\_ICU TOT\_INTUBATED ACTIVE\_CASES FATAL\_CASES RESOLVED\_CASES TOTAL\_CASES  
## 1 18 7 111 50 2441 2602  
## 2 24 17 87 20 2773 2880  
## 3 6 1 30 56 831 917  
## 4 4 1 19 21 532 572  
## 5 1 0 24 2 539 565  
## 6 16 13 62 17 1247 1326  
## INFECT\_CLOSE\_CONT INFECT\_OB\_CONGR INFECT\_OB\_HEALTH INFECT\_OB\_OTHER  
## 1 152 8 214 122  
## 2 186 15 82 109  
## 3 59 25 180 23  
## 4 42 2 89 23  
## 5 40 8 20 13  
## 6 90 17 39 54  
## INFECT\_NO\_INFO INFECT\_COMMUNITY INFECT\_TRAVEL INFECT\_PENDING P\_FATALITY\_RATE  
## 1 865 693 30 1 0.019215988  
## 2 1068 773 20 4 0.006944444  
## 3 271 190 4 0 0.061068702  
## 4 187 133 4 3 0.036713287  
## 5 193 161 7 0 0.003539823  
## 6 409 412 6 3 0.012820513  
## P\_Hospital\_RATE P\_ICU\_RATE P\_INFECT\_CLOSE\_CONT P\_INFECT\_OB\_CONGR  
## 1 0.04919293 0.006917756 0.05841660 0.003074558  
## 2 0.04409722 0.008333333 0.06458333 0.005208333  
## 3 0.05888768 0.006543075 0.06434024 0.027262814  
## 4 0.04720280 0.006993007 0.07342657 0.003496503  
## 5 0.04247788 0.001769912 0.07079646 0.014159292  
## 6 0.05354449 0.012066365 0.06787330 0.012820513  
## P\_INFECT\_OB\_HEALTH P\_INFECT\_OB\_OTHER P\_INFECT\_NO\_INFO P\_INFECT\_COMMUNITY  
## 1 0.08224443 0.04688701 0.3324366 0.2663336  
## 2 0.02847222 0.03784722 0.3708333 0.2684028  
## 3 0.19629226 0.02508179 0.2955289 0.2071974  
## 4 0.15559441 0.04020979 0.3269231 0.2325175  
## 5 0.03539823 0.02300885 0.3415929 0.2849558  
## 6 0.02941176 0.04072398 0.3084465 0.3107089  
## P\_INFECT\_TRAVEL P\_INFECT\_PENDING NBH\_NIA\_IND NBH\_SH\_UNITS NBH\_SH\_RGI  
## 1 0.011529593 0.0003843198 0 950 411  
## 2 0.006944444 0.0013888889 1 1288 1181  
## 3 0.004362050 0.0000000000 1 372 180  
## 4 0.006993007 0.0052447552 0 308 299  
## 5 0.012389380 0.0000000000 1 358 358  
## 6 0.004524887 0.0022624434 1 553 401  
## AGE1\_0\_to\_14 AGE1\_15\_to\_24 AGE1\_25\_to\_54 AGE1\_55\_to\_64 AGE1\_65\_to\_84  
## 1 5,060 5,445 13,845 3,990 4,980  
## 2 7,090 5,240 13,615 3,475 3,560  
## 3 1,730 1,410 4,160 1,195 1,880  
## 4 1,640 1,355 4,300 1,520 1,730  
## 5 1,805 1,440 3,700 1,255 1,275  
## 6 4,240 3,020 8,635 2,550 3,585  
## AGE1\_85\_up COMM\_Vehicle\_Driver COMM\_Vehicle\_PASs COMM\_Public\_Transit  
## 1 615 9,445 1,140 4,380  
## 2 300 6,965 930 4,110  
## 3 350 2,860 320 1,030  
## 4 300 2,930 235 1,345  
## 5 145 2,350 250 1,330  
## 6 575 5,335 450 2,665  
## COMM\_Walk COMM\_Bicycle COMM\_Other COMM\_Tot EDU\_Tot EDU\_None EDU\_HS\_Lower  
## 1 425 60 95 15,575 17,695 2,550 7355  
## 2 385 35 170 12,600 17,025 4,205 9080  
## 3 110 15 20 4,345 5,365 930 2310  
## 4 150 25 55 4,730 5,755 850 2530  
## 5 70 20 30 4,045 4,925 775 2445  
## 6 270 45 65 8,835 11,120 1,720 4895  
## EDU\_High\_School EDU\_College\_Trades EDU\_University DWEL\_SD\_House DWEL\_Apart  
## 1 4,805 4705 5600 3,870 2,265  
## 2 4,875 3800 4135 1,655 6,165  
## 3 1,380 1480 1570 1,575 680  
## 4 1,680 1730 1505 1,775 460  
## 5 1,670 1420 1060 990 1,375  
## 6 3,175 2905 3315 2,175 4,760  
## DWEL\_Attached DWEL\_Movable INCHH\_Under\_25000 INCHH\_25000\_50000  
## 1 4,150 0 1195 2130  
## 2 2,065 0 1490 2875  
## 3 1,020 0 410 815  
## 4 1,610 0 670 850  
## 5 845 0 445 865  
## 6 850 0 1060 2025  
## INCHH\_50000\_99999 INCHH\_Over\_100000 INCHH\_25000\_99999 INCIND\_Pop\_with\_amount  
## 1 3870 3085 6000 26,275  
## 2 3810 1700 6685 23,845  
## 3 1140 935 1955 7,970  
## 4 1295 1035 2145 8,245  
## 5 1180 720 2045 7,170  
## 6 2815 1875 4840 16,765  
## INCIND\_Population\_15up P\_INCIND\_Total\_Avg INCIND\_pop\_low\_inc  
## 1 27,840 31,771 4,550  
## 2 25,745 26,548 7,140  
## 3 8,390 32,815 1,485  
## 4 8,720 34,418 1,640  
## 5 7,640 32,012 1,695  
## 6 17,745 36,674 4,340  
## P\_INCIND\_pop\_low\_inc HH\_1\_Person HH\_2\_Persons HH\_3\_Persons HH\_4\_Persons  
## 1 14 1,800 2,430 2,020 1,985  
## 2 22 1,510 2,045 2,010 2,075  
## 3 15 570 880 645 565  
## 4 16 985 1,045 700 655  
## 5 18 635 885 665 535  
## 6 20 1,760 2,285 1,400 1,235  
## HH\_5\_Persons P\_HH\_Avg\_Size HH\_Tot IMM\_Before\_1981 IMM\_1981\_2000 IMM\_2001\_2016  
## 1 2,050 3 10,280 3,790 7860 7585  
## 2 2,215 3 9,880 2,465 7205 11815  
## 3 615 3 3,280 1,175 2075 2210  
## 4 450 3 3,845 1,320 1895 1755  
## 5 500 3 3,220 1,100 2010 1620  
## 6 1,110 3 7,785 2,835 3650 4950  
## IMM\_Non\_Imm IMM\_Non\_Perm\_res IMM\_No IMM\_Yes IND\_Cat\_11 IND\_Cat\_21 IND\_Cat\_22  
## 1 12,285 1,370 13655 19235 20 20 45  
## 2 10,425 900 11325 21485 40 10 15  
## 3 4,470 190 4660 5460 0 10 20  
## 4 5,255 130 5385 4970 25 0 20  
## 5 4,600 120 4720 4730 0 0 0  
## 6 10,110 450 10560 11435 10 15 35  
## IND\_Cat\_23 IND\_Cat\_31 IND\_Cat\_41 IND\_Cat\_44 IND\_Cat\_48 IND\_Cat\_51 IND\_Cat\_52  
## 1 1,025 2,835 675 2,020 1,695 400 755  
## 2 940 2,700 610 1,580 1,240 310 505  
## 3 455 690 215 640 465 120 200  
## 4 430 610 230 655 450 95 220  
## 5 300 510 205 540 420 110 200  
## 6 790 960 375 1,050 880 235 475  
## IND\_Cat\_53 IND\_Cat\_54 IND\_Cat\_55 IND\_Cat\_56 IND\_Cat\_61 IND\_Cat\_62 IND\_Cat\_71  
## 1 295 965 35 1,285 875 1,665 335  
## 2 185 560 30 1,105 545 1,360 215  
## 3 75 275 0 300 245 410 70  
## 4 105 300 0 470 335 535 90  
## 5 90 240 20 340 260 500 50  
## 6 185 670 30 790 590 1,015 220  
## IND\_Cat\_72 IND\_Cat\_81 IND\_Cat\_91 IND\_Tot IND\_Ess\_Yes IND\_Ess\_NO OCC\_NA  
## 1 1,195 710 460 17,295 12430 4880 565  
## 2 1,155 585 295 13,955 10665 3320 815  
## 3 325 180 145 4,860 3465 1375 210  
## 4 290 210 210 5,265 3650 1630 215  
## 5 280 190 145 4,435 3080 1320 205  
## 6 800 430 285 9,830 6715 3125 435  
## OCC\_Cat\_0 OCC\_Cat\_1 OCC\_Cat\_2 OCC\_Cat\_3 OCC\_Cat\_4 OCC\_Cat\_5 OCC\_Cat\_6  
## 1 915 2,710 1,165 1,055 1,295 290 4,740  
## 2 590 1,800 670 840 1,035 185 4,000  
## 3 350 725 210 265 395 110 1,335  
## 4 370 930 280 230 555 125 1,465  
## 5 285 700 225 290 360 85 1,330  
## 6 825 1,660 600 580 950 245 2,775  
## OCC\_Cat\_7 OCC\_Cat\_8 OCC\_Cat\_9 OCC\_Tot OCC\_Ess\_No OCC\_Ess\_Yes VM\_Black  
## 1 2,940 90 2,065 17,845 6375 11455 6,260  
## 2 2,605 85 2,120 14,765 4280 10465 8,390  
## 3 965 35 475 5,060 1790 3285 1,645  
## 4 835 80 390 5,480 2260 3215 1,470  
## 5 755 30 375 4,635 1655 2985 2,930  
## 6 1,595 40 565 10,265 4280 5990 5,340  
## VM\_South\_ASian VM\_EASt\_ASian VM\_SOUTheASt\_ASian VM\_West\_ASian  
## 1 13,920 590 2105 540  
## 2 11,095 365 1405 4065  
## 3 2,555 165 425 630  
## 4 1,420 210 685 325  
## 5 840 225 835 325  
## 6 3,630 655 955 810  
## VM\_Latin\_American VM\_Multiple VM\_Not\_Vismin VM\_Other VM\_Mult\_Oth VM\_No VM\_Yes  
## 1 1,300 685 5,940 1,535 1370 5,940 26935  
## 2 1,085 610 4,375 1,445 1220 4,375 28460  
## 3 660 160 3,615 285 320 3,615 6525  
## 4 850 170 4,990 235 340 4,990 5365  
## 5 765 245 3,085 210 490 3,085 6375  
## 6 980 425 8,835 325 850 8,835 13120  
## POP\_population POP\_land\_area POP\_density AGE\_00\_to\_19 AGE\_20\_to\_39  
## 1 33,312 30 1,117 7380 10285  
## 2 32,954 5 7,291 9675 9510  
## 3 10,360 3 3,130 2370 2820  
## 4 10,529 2 4,229 2275 2760  
## 5 9,456 3 3,306 2490 2480  
## 6 22,000 5 4,356 5740 5710  
## AGE\_40\_to\_64 AGE\_65\_UP P\_AGE1\_0\_to\_14 P\_AGE1\_15\_to\_24 P\_AGE1\_25\_to\_54  
## 1 10645 5005 0.1518972 0.1634546 0.4156160  
## 2 10265 3560 0.2151484 0.1590095 0.4131517  
## 3 3285 1885 0.1669884 0.1361004 0.4015444  
## 4 3775 1725 0.1557603 0.1286922 0.4083959  
## 5 3210 1290 0.1908841 0.1522843 0.3912860  
## 6 6970 3580 0.1927273 0.1372727 0.3925000  
## P\_AGE1\_55\_to\_64 P\_AGE1\_65\_to\_84 P\_AGE1\_85\_UP P\_AGE\_00\_to\_19 P\_AGE\_20\_to\_39  
## 1 0.1197767 0.1494957 0.018461816 0.2215418 0.3087476  
## 2 0.1054500 0.1080294 0.009103599 0.2935911 0.2885841  
## 3 0.1153475 0.1814672 0.033783784 0.2287645 0.2722008  
## 4 0.1443632 0.1643081 0.028492734 0.2160699 0.2621332  
## 5 0.1327200 0.1348350 0.015334179 0.2633249 0.2622673  
## 6 0.1159091 0.1629545 0.026136364 0.2609091 0.2595455  
## P\_AGE\_40\_to\_64 P\_AGE\_65\_up P\_COMM\_Bicycle P\_COMM\_Other P\_COMM\_Public\_Transit  
## 1 0.3195545 0.1502462 0.003852327 0.006099518 0.2812199  
## 2 0.3114948 0.1080294 0.002777778 0.013492064 0.3261905  
## 3 0.3170849 0.1819498 0.003452244 0.004602992 0.2370541  
## 4 0.3585336 0.1638332 0.005285412 0.011627907 0.2843552  
## 5 0.3394670 0.1364213 0.004944376 0.007416564 0.3288010  
## 6 0.3168182 0.1627273 0.005093379 0.007357102 0.3016412  
## P\_COMM\_Vehicle\_Driver P\_COMM\_Vehicle\_PASs P\_COMM\_Walk P\_DWEL\_Apart  
## 1 0.6064205 0.07319422 0.02728732 0.2203307  
## 2 0.5527778 0.07380952 0.03055556 0.6239879  
## 3 0.6582278 0.07364787 0.02531646 0.2073171  
## 4 0.6194503 0.04968288 0.03171247 0.1196359  
## 5 0.5809642 0.06180470 0.01730532 0.4270186  
## 6 0.6038483 0.05093379 0.03056027 0.6114322  
## P\_DWEL\_Attached P\_DWEL\_Movable P\_DWEL\_SD\_House P\_EDU\_College\_Trades  
## 1 0.4036965 0 0.3764591 0.2658943  
## 2 0.2090081 0 0.1675101 0.2232012  
## 3 0.3109756 0 0.4801829 0.2758621  
## 4 0.4187256 0 0.4616385 0.3006082  
## 5 0.2624224 0 0.3074534 0.2883249  
## 6 0.1091843 0 0.2793834 0.2612410  
## P\_EDU\_High\_School P\_EDU\_HS\_Lower P\_EDU\_None P\_EDU\_University P\_HH\_1\_Person  
## 1 0.2715456 0.4156541 0.1441085 0.3164736 0.1750973  
## 2 0.2863436 0.5333333 0.2469897 0.2428781 0.1528340  
## 3 0.2572227 0.4305685 0.1733458 0.2926375 0.1737805  
## 4 0.2919201 0.4396177 0.1476977 0.2615117 0.2561769  
## 5 0.3390863 0.4964467 0.1573604 0.2152284 0.1972050  
## 6 0.2855216 0.4401978 0.1546763 0.2981115 0.2260758  
## P\_HH\_2\_Persons P\_HH\_3\_Persons P\_HH\_4\_Persons P\_HH\_5\_Persons P\_IMM\_1981\_2000  
## 1 0.2363813 0.1964981 0.1930934 0.1994163 0.2359510  
## 2 0.2069838 0.2034413 0.2100202 0.2241903 0.2186381  
## 3 0.2682927 0.1966463 0.1722561 0.1875000 0.2002896  
## 4 0.2717815 0.1820546 0.1703511 0.1170351 0.1799791  
## 5 0.2748447 0.2065217 0.1661491 0.1552795 0.2125635  
## 6 0.2935132 0.1798330 0.1586384 0.1425819 0.1659091  
## P\_IMM\_2001\_2016 P\_IMM\_Before\_1981 P\_IMM\_Non\_Imm P\_IMM\_1Non\_Perm\_Res P\_IMM\_Yes  
## 1 0.2276957 0.11377281 0.3687860 0.04112632 0.5774195  
## 2 0.3585301 0.07480124 0.3163501 0.02731080 0.6519694  
## 3 0.2133205 0.11341699 0.4314672 0.01833977 0.5270270  
## 4 0.1666825 0.12536803 0.4990977 0.01234685 0.4720296  
## 5 0.1713198 0.11632826 0.4864636 0.01269036 0.5002115  
## 6 0.2250000 0.12886364 0.4595455 0.02045455 0.5197727  
## P\_IMM\_No P\_INC\_25000\_50000 P\_INC\_50000\_99999 P\_INC\_Over\_100000  
## 1 0.4099123 0.2071984 0.3764591 0.3000973  
## 2 0.3436609 0.2909919 0.3856275 0.1720648  
## 3 0.4498069 0.2484756 0.3475610 0.2850610  
## 4 0.5114446 0.2210663 0.3368010 0.2691808  
## 5 0.4991540 0.2686335 0.3664596 0.2236025  
## 6 0.4800000 0.2601156 0.3615928 0.2408478  
## P\_INC\_Under\_25000 P\_INC\_25000\_99999 P\_IND\_Cat\_11 P\_IND\_Cat\_21 P\_IND\_Cat\_22  
## 1 0.1162451 0.5836576 0.001156404 0.001156404 0.002601908  
## 2 0.1508097 0.6766194 0.002866356 0.000716589 0.001074884  
## 3 0.1250000 0.5960366 0.000000000 0.002057613 0.004115226  
## 4 0.1742523 0.5578674 0.004748338 0.000000000 0.003798670  
## 5 0.1381988 0.6350932 0.000000000 0.000000000 0.000000000  
## 6 0.1361593 0.6217084 0.001017294 0.001525941 0.003560529  
## P\_IND\_Cat\_23 P\_IND\_Cat\_31 P\_IND\_Cat\_41 P\_IND\_Cat\_44 P\_IND\_Cat\_48 P\_IND\_Cat\_51  
## 1 0.05926568 0.16392021 0.03902862 0.1167968 0.09800520 0.02312807  
## 2 0.06735937 0.19347904 0.04371193 0.1132211 0.08885704 0.02221426  
## 3 0.09362140 0.14197531 0.04423868 0.1316872 0.09567901 0.02469136  
## 4 0.08167141 0.11585945 0.04368471 0.1244065 0.08547009 0.01804368  
## 5 0.06764374 0.11499436 0.04622322 0.1217587 0.09470124 0.02480271  
## 6 0.08036623 0.09766022 0.03814852 0.1068159 0.08952187 0.02390641  
## P\_IND\_Cat\_52 P\_IND\_Cat\_53 P\_IND\_Cat\_54 P\_IND\_Cat\_55 P\_IND\_Cat\_56 P\_IND\_Cat\_61  
## 1 0.04365424 0.01705695 0.05579647 0.002023706 0.07429893 0.05059266  
## 2 0.03618775 0.01325690 0.04012899 0.002149767 0.07918309 0.03905410  
## 3 0.04115226 0.01543210 0.05658436 0.000000000 0.06172840 0.05041152  
## 4 0.04178538 0.01994302 0.05698006 0.000000000 0.08926876 0.06362773  
## 5 0.04509583 0.02029312 0.05411499 0.004509583 0.07666291 0.05862458  
## 6 0.04832146 0.01881994 0.06815870 0.003051882 0.08036623 0.06002035  
## P\_IND\_Cat\_62 P\_IND\_Cat\_71 P\_IND\_Cat\_72 P\_IND\_Cat\_81 P\_IND\_Cat\_91  
## 1 0.09627060 0.01936976 0.06909511 0.04105233 0.02659728  
## 2 0.09745611 0.01540666 0.08276603 0.04192046 0.02113938  
## 3 0.08436214 0.01440329 0.06687243 0.03703704 0.02983539  
## 4 0.10161443 0.01709402 0.05508072 0.03988604 0.03988604  
## 5 0.11273957 0.01127396 0.06313416 0.04284104 0.03269448  
## 6 0.10325534 0.02238047 0.08138352 0.04374364 0.02899288  
## P\_IND\_Ess\_Yes P\_IND\_Ess\_No P\_NBH\_SH\_UNITS NIA\_IND P\_OCC\_0 P\_OCC\_1  
## 1 0.7187048 0.2821625 0.010584840 0 0.05127487 0.1518633  
## 2 0.7642422 0.2379076 0.014350815 1 0.03995936 0.1219099  
## 3 0.7129630 0.2829218 0.004144801 1 0.06916996 0.1432806  
## 4 0.6932574 0.3095916 0.003431717 0 0.06751825 0.1697080  
## 5 0.6944758 0.2976325 0.003988813 1 0.06148867 0.1510248  
## 6 0.6831129 0.3179044 0.006161491 1 0.08037019 0.1617146  
## P\_OCC\_2 P\_OCC\_3 P\_OCC\_4 P\_OCC\_5 P\_OCC\_6 P\_OCC\_7 P\_OCC\_8  
## 1 0.06528439 0.05912020 0.07256935 0.01625105 0.2656206 0.1647520 0.005043430  
## 2 0.04537758 0.05689130 0.07009821 0.01252963 0.2709109 0.1764307 0.005756857  
## 3 0.04150198 0.05237154 0.07806324 0.02173913 0.2638340 0.1907115 0.006916996  
## 4 0.05109489 0.04197080 0.10127737 0.02281022 0.2673358 0.1523723 0.014598540  
## 5 0.04854369 0.06256742 0.07766990 0.01833873 0.2869471 0.1628910 0.006472492  
## 6 0.05845105 0.05650268 0.09254749 0.02386751 0.2703361 0.1553824 0.003896737  
## P\_OCC\_9 P\_OCC\_NA P\_OCC\_Ess\_Yes P\_OCC\_Ess\_No P\_VM\_Black P\_VM\_EASt\_ASian  
## 1 0.11571869 0.03166153 0.6419165 0.3572429 0.1879203 0.01771134  
## 2 0.14358280 0.05519810 0.7087707 0.2898747 0.2545973 0.01107605  
## 3 0.09387352 0.04150198 0.6492095 0.3537549 0.1587838 0.01592664  
## 4 0.07116788 0.03923358 0.5866788 0.4124088 0.1396144 0.01994491  
## 5 0.08090615 0.04422869 0.6440129 0.3570658 0.3098562 0.02379442  
## 6 0.05504140 0.04237701 0.5835363 0.4169508 0.2427273 0.02977273  
## P\_VM\_Latin\_American P\_VM\_Multiple P\_VM\_Not\_Vismin P\_VM\_Other P\_VM\_SOUTh\_ASian  
## 1 0.03902498 0.02056316 0.1783141 0.04607949 0.41786744  
## 2 0.03292468 0.01851065 0.1327608 0.04384900 0.33668143  
## 3 0.06370656 0.01544402 0.3489382 0.02750965 0.24662162  
## 4 0.08072941 0.01614588 0.4739291 0.02231931 0.13486561  
## 5 0.08090102 0.02590948 0.3262479 0.02220812 0.08883249  
## 6 0.04454545 0.01931818 0.4015909 0.01477273 0.16500000  
## P\_VM\_SOUTheASt\_ASian P\_VM\_West\_ASian P\_VM\_No P\_VM\_Yes P\_VM\_Mult\_Oth  
## 1 0.06319044 0.01621037 0.1783141 0.8085675 0.04112632  
## 2 0.04263519 0.12335377 0.1327608 0.8636281 0.03702130  
## 3 0.04102317 0.06081081 0.3489382 0.6298263 0.03088803  
## 4 0.06505841 0.03086713 0.4739291 0.5095451 0.03229177  
## 5 0.08830372 0.03436971 0.3262479 0.6741751 0.05181895  
## 6 0.04340909 0.03681818 0.4015909 0.5963636 0.03863636  
## INFECTION\_RATE  
## 1 0.078  
## 2 0.087  
## 3 0.089  
## 4 0.054  
## 5 0.060  
## 6 0.060

summary(covid\_dat)

## NEIGHBOURHOOD\_ID neighbourhood\_name TOT\_HOSPITALIZED  
## Min. : 1.00 Agincourt North : 1 Min. : 2.00   
## 1st Qu.: 35.00 Agincourt South-Malvern West: 1 1st Qu.: 19.00   
## Median : 69.00 Alderwood : 1 Median : 33.00   
## Mean : 69.77 Annex : 1 Mean : 44.66   
## 3rd Qu.:104.00 Banbury-Don Mills : 1 3rd Qu.: 60.00   
## Max. :140.00 Bathurst Manor : 1 Max. :181.00   
## (Other) :131   
## TOT\_IN\_ICU TOT\_INTUBATED ACTIVE\_CASES FATAL\_CASES   
## Min. : 0.000 Min. : 0.000 Min. : 0.00 Min. : 0.00   
## 1st Qu.: 3.000 1st Qu.: 1.000 1st Qu.: 17.00 1st Qu.: 6.00   
## Median : 6.000 Median : 3.000 Median : 26.00 Median :11.00   
## Mean : 8.015 Mean : 4.818 Mean : 36.76 Mean :19.27   
## 3rd Qu.:12.000 3rd Qu.: 7.000 3rd Qu.: 45.00 3rd Qu.:25.00   
## Max. :40.000 Max. :29.000 Max. :167.00 Max. :96.00   
##   
## RESOLVED\_CASES TOTAL\_CASES INFECT\_CLOSE\_CONT INFECT\_OB\_CONGR   
## Min. : 101.0 Min. : 109.0 Min. : 6.00 Min. : 0.000   
## 1st Qu.: 280.0 1st Qu.: 320.0 1st Qu.: 22.00 1st Qu.: 1.000   
## Median : 504.0 Median : 545.0 Median : 37.00 Median : 3.000   
## Mean : 692.6 Mean : 748.6 Mean : 47.65 Mean : 9.438   
## 3rd Qu.: 862.0 3rd Qu.: 944.0 3rd Qu.: 62.00 3rd Qu.: 9.000   
## Max. :2773.0 Max. :2941.0 Max. :233.00 Max. :180.000   
##   
## INFECT\_OB\_HEALTH INFECT\_OB\_OTHER INFECT\_NO\_INFO INFECT\_COMMUNITY  
## Min. : 3.00 Min. : 3.00 Min. : 28.0 Min. : 20.0   
## 1st Qu.: 21.00 1st Qu.: 11.00 1st Qu.: 96.0 1st Qu.: 85.0   
## Median : 46.00 Median : 21.00 Median : 168.0 Median :147.0   
## Mean : 78.12 Mean : 31.78 Mean : 239.8 Mean :198.6   
## 3rd Qu.:113.00 3rd Qu.: 43.00 3rd Qu.: 316.0 3rd Qu.:259.0   
## Max. :313.00 Max. :152.00 Max. :1068.0 Max. :804.0   
##   
## INFECT\_TRAVEL INFECT\_PENDING P\_FATALITY\_RATE P\_Hospital\_RATE   
## Min. : 0.000 Min. : 0.000 Min. :0.00000 Min. :0.01316   
## 1st Qu.: 4.000 1st Qu.: 0.000 1st Qu.:0.01115 1st Qu.:0.04921   
## Median : 6.000 Median : 1.000 Median :0.01964 Median :0.05874   
## Mean : 8.124 Mean : 1.672 Mean :0.02757 Mean :0.06219   
## 3rd Qu.:10.000 3rd Qu.: 2.000 3rd Qu.:0.03739 3rd Qu.:0.07391   
## Max. :54.000 Max. :13.000 Max. :0.14706 Max. :0.12500   
##   
## P\_ICU\_RATE P\_INFECT\_CLOSE\_CONT P\_INFECT\_OB\_CONGR P\_INFECT\_OB\_HEALTH  
## Min. :0.000000 Min. :0.02299 Min. :0.000000 Min. :0.01832   
## 1st Qu.:0.006918 1st Qu.:0.05412 1st Qu.:0.002695 1st Qu.:0.04404   
## Median :0.010204 Median :0.06458 Median :0.006122 Median :0.07866   
## Mean :0.010363 Mean :0.06746 Mean :0.011425 Mean :0.11525   
## 3rd Qu.:0.013752 3rd Qu.:0.07931 3rd Qu.:0.012766 3rd Qu.:0.15385   
## Max. :0.024457 Max. :0.14762 Max. :0.123882 Max. :0.41036   
##   
## P\_INFECT\_OB\_OTHER P\_INFECT\_NO\_INFO P\_INFECT\_COMMUNITY P\_INFECT\_TRAVEL   
## Min. :0.01182 Min. :0.1978 Min. :0.1250 Min. :0.000000   
## 1st Qu.:0.02981 1st Qu.:0.2904 1st Qu.:0.2395 1st Qu.:0.005856   
## Median :0.03984 Median :0.3206 Median :0.2715 Median :0.010348   
## Mean :0.04246 Mean :0.3153 Mean :0.2677 Mean :0.015248   
## 3rd Qu.:0.05027 3rd Qu.:0.3436 3rd Qu.:0.2941 3rd Qu.:0.020548   
## Max. :0.09572 Max. :0.4106 Max. :0.4312 Max. :0.077419   
##   
## P\_INFECT\_PENDING NBH\_NIA\_IND NBH\_SH\_UNITS NBH\_SH\_RGI   
## Min. :0.000000 Min. :0.0000 Min. : 0.0 Min. : 0.0   
## 1st Qu.:0.000000 1st Qu.:0.0000 1st Qu.: 166.0 1st Qu.: 94.0   
## Median :0.001614 Median :0.0000 Median : 468.0 Median : 328.0   
## Mean :0.002226 Mean :0.2117 Mean : 620.2 Mean : 464.8   
## 3rd Qu.:0.003534 3rd Qu.:0.0000 3rd Qu.: 879.0 3rd Qu.: 630.0   
## Max. :0.012766 Max. :1.0000 Max. :3324.0 Max. :2362.0   
##   
## AGE1\_0\_to\_14 AGE1\_15\_to\_24 AGE1\_25\_to\_54 AGE1\_55\_to\_64 AGE1\_65\_to\_84  
## 1,695 : 3 1,035 : 2 3,790 : 2 1,760 : 3 1,325 : 2   
## 1,470 : 2 1,065 : 2 5,860 : 2 2,020 : 3 1,515 : 2   
## 1,610 : 2 1,305 : 2 7,470 : 2 1,195 : 2 1,785 : 2   
## 1,675 : 2 1,355 : 2 10,310 : 1 1,325 : 2 1,865 : 2   
## 1,745 : 2 1,465 : 2 10,350 : 1 1,390 : 2 2,015 : 2   
## 1,960 : 2 2,175 : 2 10,410 : 1 1,435 : 2 2,160 : 2   
## (Other):124 (Other):125 (Other):128 (Other):123 (Other):125   
## AGE1\_85\_up COMM\_Vehicle\_Driver COMM\_Vehicle\_PASs COMM\_Public\_Transit  
## 140 : 4 1,940 : 2 275 : 5 2,010 : 2   
## 195 : 3 2,070 : 2 165 : 4 2,245 : 2   
## 255 : 3 2,860 : 2 175 : 4 2,665 : 2   
## 260 : 3 3,065 : 2 195 : 4 2,720 : 2   
## 265 : 3 3,705 : 2 355 : 4 2,845 : 2   
## 300 : 3 3,825 : 2 120 : 3 3,505 : 2   
## (Other):118 (Other):125 (Other):113 (Other):125   
## COMM\_Walk COMM\_Bicycle COMM\_Other COMM\_Tot EDU\_Tot   
## 180 : 5 10 :10 Min. : 15.0 4,345 : 2 10,645 : 2   
## 245 : 4 20 :10 1st Qu.: 55.0 4,730 : 2 14,260 : 2   
## 340 : 4 35 : 9 Median : 85.0 4,980 : 2 15,455 : 2   
## 110 : 3 55 : 6 Mean : 96.2 7,450 : 2 5,925 : 2   
## 270 : 3 30 : 5 3rd Qu.:115.0 8,075 : 2 9,070 : 2   
## 280 : 3 45 : 5 Max. :610.0 10,050 : 1 10,335 : 1   
## (Other):115 (Other):92 (Other):126 (Other):126   
## EDU\_None EDU\_HS\_Lower EDU\_High\_School EDU\_College\_Trades  
## 385 : 3 Min. : 505 3,410 : 3 Min. : 365   
## 1,080 : 2 1st Qu.: 1730 1,255 : 2 1st Qu.:1495   
## 1,105 : 2 Median : 2875 1,265 : 2 Median :2010   
## 1,720 : 2 Mean : 3401 1,405 : 2 Mean :2447   
## 155 : 2 3rd Qu.: 4370 1,475 : 2 3rd Qu.:2905   
## 255 : 2 Max. :11545 1,520 : 2 Max. :7050   
## (Other):124 (Other):124   
## EDU\_University DWEL\_SD\_House DWEL\_Apart DWEL\_Attached DWEL\_Movable   
## Min. : 645 5 : 3 0 : 4 1,165 : 2 Min. : 0.0000   
## 1st Qu.: 2875 0 : 2 1,335 : 2 1,325 : 2 1st Qu.: 0.0000   
## Median : 4425 1,635 : 2 1,965 : 2 1,610 : 2 Median : 0.0000   
## Mean : 5273 1,680 : 2 1,995 : 2 1,790 : 2 Mean : 0.6934   
## 3rd Qu.: 6120 2,025 : 2 2,025 : 2 1,795 : 2 3rd Qu.: 0.0000   
## Max. :36875 2,175 : 2 2,265 : 2 2,030 : 2 Max. :20.0000   
## (Other):124 (Other):123 (Other):125   
## INCHH\_Under\_25000 INCHH\_25000\_50000 INCHH\_50000\_99999 INCHH\_Over\_100000  
## Min. : 175 Min. : 190 Min. : 470 Min. : 420   
## 1st Qu.: 705 1st Qu.: 945 1st Qu.: 1390 1st Qu.: 1380   
## Median :1145 Median :1350 Median : 1915 Median : 2105   
## Mean :1373 Mean :1656 Mean : 2449 Mean : 2469   
## 3rd Qu.:1645 3rd Qu.:2150 3rd Qu.: 2935 3rd Qu.: 2915   
## Max. :6370 Max. :5910 Max. :13520 Max. :14975   
##   
## INCHH\_25000\_99999 INCIND\_Pop\_with\_amount INCIND\_Population\_15up  
## Min. : 660 10,805 : 2 12,580 : 2   
## 1st Qu.: 2350 11,780 : 2 18,585 : 2   
## Median : 3330 13,510 : 2 6,280 : 2   
## Mean : 4105 21,540 : 2 10,000 : 1   
## 3rd Qu.: 5205 7,970 : 2 10,105 : 1   
## Max. :19430 10,060 : 1 10,195 : 1   
## (Other):126 (Other):128   
## P\_INCIND\_Total\_Avg INCIND\_pop\_low\_inc P\_INCIND\_pop\_low\_inc HH\_1\_Person   
## 100,516: 1 1,060 : 2 Min. : 4.00 1,785 : 4   
## 101,551: 1 1,640 : 2 1st Qu.:13.00 1,105 : 2   
## 111,730: 1 1,910 : 2 Median :16.00 1,195 : 2   
## 112,766: 1 1,945 : 2 Mean :16.77 1,285 : 2   
## 114,174: 1 2,055 : 2 3rd Qu.:20.00 1,660 : 2   
## 123,077: 1 2,100 : 2 Max. :36.00 1,955 : 2   
## (Other):131 (Other):125 (Other):123   
## HH\_2\_Persons HH\_3\_Persons HH\_4\_Persons HH\_5\_Persons P\_HH\_Avg\_Size   
## 1,135 : 2 550 : 3 970 : 6 370 : 4 Min. :2.000   
## 1,150 : 2 710 : 3 755 : 3 590 : 4 1st Qu.:2.000   
## 1,345 : 2 720 : 3 820 : 3 265 : 3 Median :3.000   
## 1,455 : 2 740 : 3 1,375 : 2 285 : 3 Mean :2.518   
## 1,620 : 2 1,005 : 2 1,515 : 2 450 : 3 3rd Qu.:3.000   
## 1,630 : 2 1,065 : 2 2,225 : 2 525 : 3 Max. :3.000   
## (Other):125 (Other):121 (Other):119 (Other):117   
## HH\_Tot IMM\_Before\_1981 IMM\_1981\_2000 IMM\_2001\_2016 IMM\_Non\_Imm   
## 5,450 : 2 815 : 3 Min. : 505 Min. : 340 10,080 : 1   
## 6,435 : 2 1,045 : 2 1st Qu.: 1525 1st Qu.: 1325 10,110 : 1   
## 6,570 : 2 1,175 : 2 Median : 2420 Median : 2880 10,180 : 1   
## 7,020 : 2 1,235 : 2 Mean : 3245 Mean : 3714 10,280 : 1   
## 10,050 : 1 1,310 : 2 3rd Qu.: 4100 3rd Qu.: 5415 10,295 : 1   
## 10,060 : 1 1,335 : 2 Max. :13030 Max. :17065 10,425 : 1   
## (Other):127 (Other):124 (Other):131   
## IMM\_Non\_Perm\_res IMM\_No IMM\_Yes IND\_Cat\_11   
## 255 : 4 Min. : 2945 Min. : 1805 Min. : 0.00   
## 390 : 3 1st Qu.: 6625 1st Qu.: 4520 1st Qu.:10.00   
## 425 : 3 Median : 8825 Median : 7485 Median :10.00   
## 470 : 3 Mean :10211 Mean : 9073 Mean :15.07   
## 1,185 : 2 3rd Qu.:12115 3rd Qu.:11435 3rd Qu.:20.00   
## 130 : 2 Max. :41810 Max. :31045 Max. :60.00   
## (Other):120   
## IND\_Cat\_21 IND\_Cat\_22 IND\_Cat\_23 IND\_Cat\_31 IND\_Cat\_41   
## Min. : 0.00 Min. : 0.00 350 : 5 610 : 4 320 : 5   
## 1st Qu.: 0.00 1st Qu.: 20.00 180 : 4 200 : 3 170 : 4   
## Median : 10.00 Median : 35.00 540 : 4 350 : 3 220 : 4   
## Mean : 14.71 Mean : 41.97 300 : 3 545 : 3 245 : 4   
## 3rd Qu.: 20.00 3rd Qu.: 55.00 325 : 3 175 : 2 300 : 4   
## Max. :185.00 Max. :250.00 425 : 3 180 : 2 155 : 3   
## (Other):115 (Other):120 (Other):113   
## IND\_Cat\_44 IND\_Cat\_48 IND\_Cat\_51 IND\_Cat\_52 IND\_Cat\_53   
## 645 : 3 215 : 4 265 : 5 375 : 4 255 : 6   
## 1,050 : 2 345 : 4 270 : 4 475 : 3 170 : 4   
## 1,215 : 2 160 : 3 310 : 4 1,020 : 2 175 : 4   
## 1,290 : 2 190 : 3 130 : 3 1,190 : 2 185 : 4   
## 1,540 : 2 370 : 3 170 : 3 1,210 : 2 270 : 4   
## 1,580 : 2 625 : 3 220 : 3 200 : 2 295 : 4   
## (Other):124 (Other):117 (Other):115 (Other):122 (Other):111   
## IND\_Cat\_54 IND\_Cat\_55 IND\_Cat\_56 IND\_Cat\_61 IND\_Cat\_62   
## 840 : 3 Min. : 0.00 200 : 3 590 : 4 730 : 6   
## 1,100 : 2 1st Qu.: 10.00 280 : 3 1,205 : 3 1,235 : 3   
## 1,120 : 2 Median : 25.00 365 : 3 260 : 3 535 : 3   
## 1,145 : 2 Mean : 29.93 410 : 3 545 : 3 915 : 3   
## 1,170 : 2 3rd Qu.: 40.00 470 : 3 925 : 3 1,140 : 2   
## 1,430 : 2 Max. :250.00 1,405 : 2 1,040 : 2 2,570 : 2   
## (Other):124 (Other):120 (Other):119 (Other):118   
## IND\_Cat\_71 IND\_Cat\_72 IND\_Cat\_81 IND\_Cat\_91 IND\_Tot   
## 115 : 6 440 : 3 275 : 4 280 : 5 4,605 : 2   
## 155 : 6 465 : 3 315 : 4 145 : 4 5,005 : 2   
## 165 : 5 1,125 : 2 350 : 4 210 : 4 5,345 : 2   
## 125 : 4 1,145 : 2 530 : 4 220 : 4 5,980 : 2   
## 190 : 4 235 : 2 295 : 3 255 : 4 8,285 : 2   
## 225 : 4 260 : 2 340 : 3 265 : 4 10,115 : 1   
## (Other):108 (Other):123 (Other):115 (Other):112 (Other):126   
## IND\_Ess\_Yes IND\_Ess\_NO OCC\_NA OCC\_Cat\_0 OCC\_Cat\_1   
## Min. : 1600 Min. : 780 130 : 5 1,020 : 4 1,180 : 3   
## 1st Qu.: 3505 1st Qu.: 2740 200 : 5 590 : 3 1,490 : 3   
## Median : 4645 Median : 4075 105 : 4 895 : 3 1,555 : 3   
## Mean : 5561 Mean : 4733 150 : 4 1,090 : 2 1,010 : 2   
## 3rd Qu.: 6980 3rd Qu.: 5155 100 : 3 1,095 : 2 1,035 : 2   
## Max. :15920 Max. :34860 145 : 3 1,105 : 2 1,120 : 2   
## (Other):113 (Other):121 (Other):122   
## OCC\_Cat\_2 OCC\_Cat\_3 OCC\_Cat\_4 OCC\_Cat\_5 OCC\_Cat\_6   
## 345 : 4 230 : 4 1,090 : 3 350 : 4 1,220 : 2   
## 400 : 3 255 : 3 950 : 3 420 : 4 1,290 : 2   
## 630 : 3 405 : 3 1,050 : 2 200 : 3 1,465 : 2   
## 1,150 : 2 410 : 3 1,180 : 2 275 : 3 1,625 : 2   
## 1,400 : 2 415 : 3 1,365 : 2 290 : 3 1,710 : 2   
## 225 : 2 565 : 3 1,385 : 2 345 : 3 2,180 : 2   
## (Other):121 (Other):118 (Other):123 (Other):117 (Other):125   
## OCC\_Cat\_7 OCC\_Cat\_8 OCC\_Cat\_9 OCC\_Tot OCC\_Ess\_No   
## 1,040 : 3 Min. : 0.00 75 : 5 10,105 : 2 Min. : 1140   
## 480 : 3 1st Qu.: 30.00 150 : 4 5,115 : 2 1st Qu.: 3440   
## 1,105 : 2 Median : 50.00 35 : 4 5,445 : 2 Median : 4915   
## 1,205 : 2 Mean : 52.92 95 : 4 8,620 : 2 Mean : 5789   
## 1,270 : 2 3rd Qu.: 70.00 155 : 3 8,890 : 2 3rd Qu.: 6375   
## 130 : 2 Max. :180.00 160 : 3 9,275 : 2 Max. :38830   
## (Other):123 (Other):114 (Other):125   
## OCC\_Ess\_Yes VM\_Black VM\_South\_ASian VM\_EASt\_ASian   
## Min. : 1170 1,285 : 2 445 : 3 Min. : 125   
## 1st Qu.: 2950 1,465 : 2 230 : 2 1st Qu.: 610   
## Median : 3945 180 : 2 245 : 2 Median : 1050   
## Mean : 4833 2,130 : 2 325 : 2 Mean : 2567   
## 3rd Qu.: 6105 205 : 2 330 : 2 3rd Qu.: 2500   
## Max. :15075 220 : 2 395 : 2 Max. :22900   
## (Other):125 (Other):124   
## VM\_SOUTheASt\_ASian VM\_West\_ASian VM\_Latin\_American VM\_Multiple   
## Min. : 95 Min. : 10.0 120 : 3 230 : 6   
## 1st Qu.: 490 1st Qu.: 145.0 140 : 3 100 : 4   
## Median : 845 Median : 320.0 400 : 3 120 : 4   
## Mean :1387 Mean : 694.6 110 : 2 170 : 4   
## 3rd Qu.:1965 3rd Qu.: 830.0 135 : 2 180 : 4   
## Max. :6600 Max. :7305.0 145 : 2 245 : 4   
## (Other):122 (Other):111   
## VM\_Not\_Vismin VM\_Other VM\_Mult\_Oth VM\_No VM\_Yes   
## 6,075 : 2 65 : 5 Min. : 70.0 6,075 : 2 Min. : 1110   
## 10,115 : 1 70 : 5 1st Qu.: 340.0 10,115 : 1 1st Qu.: 3630   
## 10,125 : 1 120 : 4 Median : 520.0 10,125 : 1 Median : 6905   
## 10,165 : 1 140 : 4 Mean : 681.4 10,165 : 1 Mean : 9907   
## 10,185 : 1 150 : 4 3rd Qu.: 880.0 10,185 : 1 3rd Qu.:12720   
## 10,570 : 1 45 : 4 Max. :2840.0 10,570 : 1 Max. :40510   
## (Other):130 (Other):111 (Other):130   
## POP\_population POP\_land\_area POP\_density AGE\_00\_to\_19 AGE\_20\_to\_39   
## 10,070 : 1 Min. : 0.000 3,614 : 2 Min. : 900 Min. : 1610   
## 10,084 : 1 1st Qu.: 2.000 5,395 : 2 1st Qu.: 2350 1st Qu.: 3370   
## 10,111 : 1 Median : 3.000 1,040 : 1 Median : 3280 Median : 4895   
## 10,360 : 1 Mean : 4.562 1,117 : 1 Mean : 3899 Mean : 6059   
## 10,529 : 1 3rd Qu.: 5.000 1,260 : 1 3rd Qu.: 4820 3rd Qu.: 7145   
## 10,554 : 1 Max. :37.000 1,570 : 1 Max. :13045 Max. :41675   
## (Other):131 (Other):129   
## AGE\_40\_to\_64 AGE\_65\_UP P\_AGE1\_0\_to\_14 P\_AGE1\_15\_to\_24   
## Min. : 1770 Min. : 740 Min. :0.0402 Min. :0.07344   
## 1st Qu.: 4125 1st Qu.:1865 1st Qu.:0.1296 1st Qu.:0.10443   
## Median : 5610 Median :2660 Median :0.1503 Median :0.12285   
## Mean : 6525 Mean :3070 Mean :0.1488 Mean :0.12280   
## 3rd Qu.: 8340 3rd Qu.:3785 3rd Qu.:0.1704 3rd Qu.:0.13743   
## Max. :16860 Max. :8990 Max. :0.2653 Max. :0.26592   
##   
## P\_AGE1\_25\_to\_54 P\_AGE1\_55\_to\_64 P\_AGE1\_65\_to\_84 P\_AGE1\_85\_UP   
## Min. :0.3335 Min. :0.06046 Min. :0.04843 Min. :0.003047   
## 1st Qu.:0.3958 1st Qu.:0.11535 1st Qu.:0.13363 1st Qu.:0.016477   
## Median :0.4262 Median :0.12527 Median :0.15454 Median :0.022817   
## Mean :0.4421 Mean :0.12570 Mean :0.16053 Mean :0.025185   
## 3rd Qu.:0.4757 3rd Qu.:0.13805 3rd Qu.:0.18324 3rd Qu.:0.031557   
## Max. :0.7479 Max. :0.16786 Max. :0.28014 Max. :0.086774   
##   
## P\_AGE\_00\_to\_19 P\_AGE\_20\_to\_39 P\_AGE\_40\_to\_64 P\_AGE\_65\_up   
## Min. :0.07244 Min. :0.1737 Min. :0.1987 Min. :0.04795   
## 1st Qu.:0.17948 1st Qu.:0.2546 1st Qu.:0.3194 1st Qu.:0.13360   
## Median :0.20573 Median :0.2767 Median :0.3449 Median :0.15426   
## Mean :0.20251 Mean :0.2975 Mean :0.3391 Mean :0.16055   
## 3rd Qu.:0.23125 3rd Qu.:0.3184 3rd Qu.:0.3600 3rd Qu.:0.18369   
## Max. :0.32381 Max. :0.6355 Max. :0.4006 Max. :0.28014   
##   
## P\_COMM\_Bicycle P\_COMM\_Other P\_COMM\_Public\_Transit  
## Min. :0.000000 Min. :0.002620 Min. :0.1801   
## 1st Qu.:0.003620 1st Qu.:0.007547 1st Qu.:0.3156   
## Median :0.008574 Median :0.010526 Median :0.3764   
## Mean :0.028435 Mean :0.010867 Mean :0.3740   
## 3rd Qu.:0.038048 3rd Qu.:0.013240 3rd Qu.:0.4303   
## Max. :0.190661 Max. :0.023307 Max. :0.6369   
##   
## P\_COMM\_Vehicle\_Driver P\_COMM\_Vehicle\_PASs P\_COMM\_Walk P\_DWEL\_Apart   
## Min. :0.1309 Min. :0.01351 Min. :0.01023 Min. :0.0000   
## 1st Qu.:0.3807 1st Qu.:0.03067 1st Qu.:0.02799 1st Qu.:0.1845   
## Median :0.4904 Median :0.04418 Median :0.03930 Median :0.3431   
## Mean :0.4720 Mean :0.04603 Mean :0.06849 Mean :0.3720   
## 3rd Qu.:0.5810 3rd Qu.:0.05673 3rd Qu.:0.06614 3rd Qu.:0.5245   
## Max. :0.7323 Max. :0.10926 Max. :0.49429 Max. :0.9818   
##   
## P\_DWEL\_Attached P\_DWEL\_Movable P\_DWEL\_SD\_House P\_EDU\_College\_Trades  
## Min. :0.01857 Min. :0.000e+00 Min. :0.0000 Min. :0.0804   
## 1st Qu.:0.18170 1st Qu.:0.000e+00 1st Qu.:0.1500 1st Qu.:0.1839   
## Median :0.30983 Median :0.000e+00 Median :0.2697 Median :0.2252   
## Mean :0.34333 Mean :9.777e-05 Mean :0.2845 Mean :0.2223   
## 3rd Qu.:0.45221 3rd Qu.:0.000e+00 3rd Qu.:0.4063 3rd Qu.:0.2667   
## Max. :0.90239 Max. :2.905e-03 Max. :0.8525 Max. :0.3128   
##   
## P\_EDU\_High\_School P\_EDU\_HS\_Lower P\_EDU\_None P\_EDU\_University  
## Min. :0.08509 Min. :0.1012 Min. :0.01083 Min. :0.1410   
## 1st Qu.:0.15146 1st Qu.:0.1940 1st Qu.:0.04258 1st Qu.:0.3446   
## Median :0.20881 Median :0.2972 Median :0.09653 Median :0.4733   
## Mean :0.20518 Mean :0.3090 Mean :0.10383 Mean :0.4687   
## 3rd Qu.:0.25946 3rd Qu.:0.4045 3rd Qu.:0.14501 3rd Qu.:0.6179   
## Max. :0.33909 Max. :0.6118 Max. :0.33440 Max. :0.8102   
##   
## P\_HH\_1\_Person P\_HH\_2\_Persons P\_HH\_3\_Persons P\_HH\_4\_Persons   
## Min. :0.1122 Min. :0.2070 Min. :0.06101 Min. :0.01754   
## 1st Qu.:0.2341 1st Qu.:0.2748 1st Qu.:0.14783 1st Qu.:0.12148   
## Median :0.2814 Median :0.2972 Median :0.17424 Median :0.15209   
## Mean :0.2989 Mean :0.2969 Mean :0.16615 Mean :0.14397   
## 3rd Qu.:0.3526 3rd Qu.:0.3175 3rd Qu.:0.19107 3rd Qu.:0.17210   
## Max. :0.5803 Max. :0.3650 Max. :0.22027 Max. :0.24191   
##   
## P\_HH\_5\_Persons P\_IMM\_1981\_2000 P\_IMM\_2001\_2016 P\_IMM\_Before\_1981  
## Min. :0.006863 Min. :0.05425 Min. :0.03667 Min. :0.03980   
## 1st Qu.:0.055670 1st Qu.:0.11055 1st Qu.:0.09836 1st Qu.:0.09151   
## Median :0.084495 Median :0.15042 Median :0.16002 Median :0.10857   
## Mean :0.094027 Mean :0.15592 Mean :0.17398 Mean :0.11197   
## 3rd Qu.:0.125618 3rd Qu.:0.20642 3rd Qu.:0.22770 3rd Qu.:0.13312   
## Max. :0.236744 Max. :0.30382 Max. :0.42827 Max. :0.19150   
##   
## P\_IMM\_Non\_Imm P\_IMM\_1Non\_Perm\_Res P\_IMM\_Yes P\_IMM\_No   
## Min. :0.2364 Min. :0.002994 Min. :0.1966 Min. :0.2772   
## 1st Qu.:0.3978 1st Qu.:0.015398 1st Qu.:0.3161 1st Qu.:0.4383   
## Median :0.4952 Median :0.026131 Median :0.4525 Median :0.5378   
## Mean :0.5126 Mean :0.030833 Mean :0.4419 Mean :0.5434   
## 3rd Qu.:0.6362 3rd Qu.:0.036323 3rd Qu.:0.5547 3rd Qu.:0.6637   
## Max. :0.7836 Max. :0.199442 Max. :0.7087 Max. :0.7987   
##   
## P\_INC\_25000\_50000 P\_INC\_50000\_99999 P\_INC\_Over\_100000 P\_INC\_Under\_25000  
## Min. :0.05864 Min. :0.1451 Min. :0.1220 Min. :0.0400   
## 1st Qu.:0.17511 1st Qu.:0.2727 1st Qu.:0.2428 1st Qu.:0.1291   
## Median :0.20720 Median :0.3115 Median :0.3001 Median :0.1620   
## Mean :0.20971 Mean :0.3021 Mean :0.3209 Mean :0.1675   
## 3rd Qu.:0.24747 3rd Qu.:0.3383 3rd Qu.:0.3842 3rd Qu.:0.2031   
## Max. :0.32725 Max. :0.3917 Max. :0.7284 Max. :0.3486   
##   
## P\_INC\_25000\_99999 P\_IND\_Cat\_11 P\_IND\_Cat\_21 P\_IND\_Cat\_22   
## Min. :0.2037 Min. :0.0000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.4587 1st Qu.:0.0005495 1st Qu.:0.000000 1st Qu.:0.002654   
## Median :0.5164 Median :0.0013643 Median :0.001193 Median :0.003929   
## Mean :0.5118 Mean :0.0015978 Mean :0.001341 Mean :0.004089   
## 3rd Qu.:0.5761 3rd Qu.:0.0024140 3rd Qu.:0.001924 3rd Qu.:0.005113   
## Max. :0.6766 Max. :0.0056625 Max. :0.011325 Max. :0.014463   
##   
## P\_IND\_Cat\_23 P\_IND\_Cat\_31 P\_IND\_Cat\_41 P\_IND\_Cat\_44   
## Min. :0.01464 Min. :0.02307 Min. :0.01731 Min. :0.05210   
## 1st Qu.:0.03846 1st Qu.:0.04196 1st Qu.:0.02716 1st Qu.:0.08192   
## Median :0.05015 Median :0.06282 Median :0.03530 Median :0.10290   
## Mean :0.05723 Mean :0.07252 Mean :0.03475 Mean :0.09962   
## 3rd Qu.:0.06764 3rd Qu.:0.09457 3rd Qu.:0.04215 3rd Qu.:0.11528   
## Max. :0.17222 Max. :0.21086 Max. :0.05853 Max. :0.13833   
##   
## P\_IND\_Cat\_48 P\_IND\_Cat\_51 P\_IND\_Cat\_52 P\_IND\_Cat\_53   
## Min. :0.01027 Min. :0.01143 Min. :0.02457 Min. :0.01326   
## 1st Qu.:0.02545 1st Qu.:0.02474 1st Qu.:0.05676 1st Qu.:0.02166   
## Median :0.03827 Median :0.03561 Median :0.07612 Median :0.02512   
## Mean :0.04144 Mean :0.04137 Mean :0.07834 Mean :0.02770   
## 3rd Qu.:0.05438 3rd Qu.:0.05429 3rd Qu.:0.09269 3rd Qu.:0.03157   
## Max. :0.09801 Max. :0.10275 Max. :0.19210 Max. :0.06162   
##   
## P\_IND\_Cat\_54 P\_IND\_Cat\_55 P\_IND\_Cat\_56 P\_IND\_Cat\_61   
## Min. :0.03167 Min. :0.000000 Min. :0.02091 Min. :0.03905   
## 1st Qu.:0.07551 1st Qu.:0.001595 1st Qu.:0.03841 1st Qu.:0.06035   
## Median :0.11157 Median :0.002538 Median :0.05237 Median :0.07664   
## Mean :0.11666 Mean :0.002789 Mean :0.05555 Mean :0.08014   
## 3rd Qu.:0.15973 3rd Qu.:0.003928 3rd Qu.:0.07140 3rd Qu.:0.09688   
## Max. :0.24342 Max. :0.009229 Max. :0.10766 Max. :0.18288   
##   
## P\_IND\_Cat\_62 P\_IND\_Cat\_71 P\_IND\_Cat\_72 P\_IND\_Cat\_81   
## Min. :0.06721 Min. :0.008991 Min. :0.02378 Min. :0.02522   
## 1st Qu.:0.08806 1st Qu.:0.015251 1st Qu.:0.05577 1st Qu.:0.04236   
## Median :0.09855 Median :0.021570 Median :0.06948 Median :0.04597   
## Mean :0.10098 Mean :0.024593 Mean :0.07300 Mean :0.04848   
## 3rd Qu.:0.11086 3rd Qu.:0.031601 3rd Qu.:0.08369 3rd Qu.:0.05222   
## Max. :0.16704 Max. :0.059446 Max. :0.15206 Max. :0.09384   
##   
## P\_IND\_Cat\_91 P\_IND\_Ess\_Yes P\_IND\_Ess\_No P\_NBH\_SH\_UNITS   
## Min. :0.01451 Min. :0.3088 Min. :0.2006 Min. :0.000000   
## 1st Qu.:0.02978 1st Qu.:0.4531 1st Qu.:0.3386 1st Qu.:0.001850   
## Median :0.03693 Median :0.5490 Median :0.4519 Median :0.005214   
## Mean :0.03776 Mean :0.5488 Mean :0.4511 Mean :0.006910   
## 3rd Qu.:0.04531 3rd Qu.:0.6589 3rd Qu.:0.5463 3rd Qu.:0.009794   
## Max. :0.06614 Max. :0.7977 Max. :0.6921 Max. :0.037036   
##   
## NIA\_IND P\_OCC\_0 P\_OCC\_1 P\_OCC\_2   
## Min. :0.0000 Min. :0.03602 Min. :0.1049 Min. :0.02613   
## 1st Qu.:0.0000 1st Qu.:0.07440 1st Qu.:0.1488 1st Qu.:0.05938   
## Median :0.0000 Median :0.10590 Median :0.1670 Median :0.07298   
## Mean :0.2117 Mean :0.11050 Mean :0.1698 Mean :0.07522   
## 3rd Qu.:0.0000 3rd Qu.:0.13952 3rd Qu.:0.1847 3rd Qu.:0.08760   
## Max. :1.0000 Max. :0.24426 Max. :0.2650 Max. :0.16270   
##   
## P\_OCC\_3 P\_OCC\_4 P\_OCC\_5 P\_OCC\_6   
## Min. :0.02611 Min. :0.05904 Min. :0.01253 Min. :0.1130   
## 1st Qu.:0.04683 1st Qu.:0.09947 1st Qu.:0.02479 1st Qu.:0.1986   
## Median :0.05375 Median :0.12321 Median :0.04107 Median :0.2283   
## Mean :0.05655 Mean :0.12996 Mean :0.05479 Mean :0.2330   
## 3rd Qu.:0.06193 3rd Qu.:0.15949 3rd Qu.:0.07600 3rd Qu.:0.2735   
## Max. :0.11967 Max. :0.22588 Max. :0.15909 Max. :0.3361   
##   
## P\_OCC\_7 P\_OCC\_8 P\_OCC\_9 P\_OCC\_NA   
## Min. :0.01000 Min. :0.000000 Min. :0.002594 Min. :0.01037   
## 1st Qu.:0.04959 1st Qu.:0.003442 1st Qu.:0.015363 1st Qu.:0.02053   
## Median :0.09137 Median :0.004973 Median :0.030271 Median :0.03005   
## Mean :0.09275 Mean :0.005490 Mean :0.040705 Mean :0.03100   
## 3rd Qu.:0.12456 3rd Qu.:0.007009 3rd Qu.:0.061105 3rd Qu.:0.03972   
## Max. :0.22409 Max. :0.016216 Max. :0.158065 Max. :0.06829   
##   
## P\_OCC\_Ess\_Yes P\_OCC\_Ess\_No P\_VM\_Black P\_VM\_EASt\_ASian   
## Min. :0.2289 Min. :0.2570 Min. :0.008629 Min. :0.01108   
## 1st Qu.:0.3498 1st Qu.:0.4335 1st Qu.:0.034650 1st Qu.:0.03933   
## Median :0.4608 Median :0.5399 Median :0.052824 Median :0.06870   
## Mean :0.4595 Mean :0.5403 Mean :0.083948 Mean :0.11572   
## 3rd Qu.:0.5665 3rd Qu.:0.6529 3rd Qu.:0.111919 3rd Qu.:0.11234   
## Max. :0.7435 Max. :0.7711 Max. :0.347550 Max. :0.72838   
##   
## P\_VM\_Latin\_American P\_VM\_Multiple P\_VM\_Not\_Vismin P\_VM\_Other   
## Min. :0.001218 Min. :0.003316 Min. :0.03387 Min. :0.000000   
## 1st Qu.:0.012904 1st Qu.:0.012106 1st Qu.:0.32635 1st Qu.:0.005077   
## Median :0.018206 Median :0.015747 Median :0.56733 Median :0.008417   
## Mean :0.029221 Mean :0.016447 Mean :0.52290 Mean :0.012305   
## 3rd Qu.:0.033038 3rd Qu.:0.019766 3rd Qu.:0.70484 3rd Qu.:0.015854   
## Max. :0.120691 Max. :0.036344 Max. :0.86938 Max. :0.073230   
##   
## P\_VM\_SOUTh\_ASian P\_VM\_SOUTheASt\_ASian P\_VM\_West\_ASian P\_VM\_No   
## Min. :0.01328 Min. :0.01025 Min. :0.001005 Min. :0.03387   
## 1st Qu.:0.03685 1st Qu.:0.03089 1st Qu.:0.010517 1st Qu.:0.32635   
## Median :0.06035 Median :0.05240 Median :0.018640 Median :0.56733   
## Mean :0.10647 Mean :0.06736 Mean :0.031066 Mean :0.52290   
## 3rd Qu.:0.13487 3rd Qu.:0.09291 3rd Qu.:0.037049 3rd Qu.:0.70484   
## Max. :0.46641 Max. :0.23255 Max. :0.153445 Max. :0.86938   
##   
## P\_VM\_Yes P\_VM\_Mult\_Oth INFECTION\_RATE   
## Min. :0.1197 Min. :0.006633 Min. :0.01000   
## 1st Qu.:0.2704 1st Qu.:0.024212 1st Qu.:0.02100   
## Median :0.4248 Median :0.031493 Median :0.03100   
## Mean :0.4625 Mean :0.032895 Mean :0.03706   
## 3rd Qu.:0.6606 3rd Qu.:0.039532 3rd Qu.:0.05100   
## Max. :0.9474 Max. :0.072687 Max. :0.08900   
##

str(covid\_dat)

## 'data.frame': 137 obs. of 226 variables:  
## $ NEIGHBOURHOOD\_ID : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ neighbourhood\_name : Factor w/ 137 levels "Agincourt North",..: 123 80 116 100 39 63 129 54 37 98 ...  
## $ TOT\_HOSPITALIZED : int 128 127 54 27 24 71 55 41 45 17 ...  
## $ TOT\_IN\_ICU : int 18 24 6 4 1 16 8 4 7 2 ...  
## $ TOT\_INTUBATED : int 7 17 1 1 0 13 3 3 3 0 ...  
## $ ACTIVE\_CASES : int 111 87 30 19 24 62 43 17 17 12 ...  
## $ FATAL\_CASES : int 50 20 56 21 2 17 8 50 18 2 ...  
## $ RESOLVED\_CASES : int 2441 2773 831 532 539 1247 828 460 474 221 ...  
## $ TOTAL\_CASES : int 2602 2880 917 572 565 1326 879 527 509 235 ...  
## $ INFECT\_CLOSE\_CONT : int 152 186 59 42 40 90 60 23 28 13 ...  
## $ INFECT\_OB\_CONGR : int 8 15 25 2 8 17 1 5 1 3 ...  
## $ INFECT\_OB\_HEALTH : int 214 82 180 89 20 39 46 192 112 10 ...  
## $ INFECT\_OB\_OTHER : int 122 109 23 23 13 54 37 19 17 13 ...  
## $ INFECT\_NO\_INFO : int 865 1068 271 187 193 409 280 123 147 78 ...  
## $ INFECT\_COMMUNITY : int 693 773 190 133 161 412 285 103 127 63 ...  
## $ INFECT\_TRAVEL : int 30 20 4 4 7 6 4 3 6 6 ...  
## $ INFECT\_PENDING : int 1 4 0 3 0 3 4 3 0 3 ...  
## $ P\_FATALITY\_RATE : num 0.01922 0.00694 0.06107 0.03671 0.00354 ...  
## $ P\_Hospital\_RATE : num 0.0492 0.0441 0.0589 0.0472 0.0425 ...  
## $ P\_ICU\_RATE : num 0.00692 0.00833 0.00654 0.00699 0.00177 ...  
## $ P\_INFECT\_CLOSE\_CONT : num 0.0584 0.0646 0.0643 0.0734 0.0708 ...  
## $ P\_INFECT\_OB\_CONGR : num 0.00307 0.00521 0.02726 0.0035 0.01416 ...  
## $ P\_INFECT\_OB\_HEALTH : num 0.0822 0.0285 0.1963 0.1556 0.0354 ...  
## $ P\_INFECT\_OB\_OTHER : num 0.0469 0.0378 0.0251 0.0402 0.023 ...  
## $ P\_INFECT\_NO\_INFO : num 0.332 0.371 0.296 0.327 0.342 ...  
## $ P\_INFECT\_COMMUNITY : num 0.266 0.268 0.207 0.233 0.285 ...  
## $ P\_INFECT\_TRAVEL : num 0.01153 0.00694 0.00436 0.00699 0.01239 ...  
## $ P\_INFECT\_PENDING : num 0.000384 0.001389 0 0.005245 0 ...  
## $ NBH\_NIA\_IND : int 0 1 1 0 1 1 0 0 0 0 ...  
## $ NBH\_SH\_UNITS : int 950 1288 372 308 358 553 762 688 0 0 ...  
## $ NBH\_SH\_RGI : int 411 1181 180 299 358 401 511 390 0 0 ...  
## $ AGE1\_0\_to\_14 : Factor w/ 126 levels "1,120","1,150",..: 113 122 30 27 36 103 89 17 46 33 ...  
## $ AGE1\_15\_to\_24 : Factor w/ 128 levels "1,035","1,040",..: 113 112 20 16 24 85 79 7 43 31 ...  
## $ AGE1\_25\_to\_54 : Factor w/ 134 levels "10,310","10,350",..: 23 22 54 56 48 121 117 50 81 51 ...  
## $ AGE1\_55\_to\_64 : Factor w/ 122 levels "1,050","1,100",..: 106 93 5 21 9 71 80 20 64 44 ...  
## $ AGE1\_65\_to\_84 : Factor w/ 126 levels "1,025","1,095",..: 105 82 29 24 5 83 104 69 74 35 ...  
## $ AGE1\_85\_up : Factor w/ 93 levels "1,000","1,040",..: 68 34 41 34 12 65 86 92 73 39 ...  
## $ COMM\_Vehicle\_Driver : Factor w/ 128 levels "1,075","1,100",..: 127 119 49 53 34 101 115 47 92 71 ...  
## $ COMM\_Vehicle\_PASs : Factor w/ 82 levels "1,060","1,140",..: 2 80 39 24 27 55 62 21 42 23 ...  
## $ COMM\_Public\_Transit : Factor w/ 130 levels "1,030","1,085",..: 101 97 1 11 10 54 46 5 34 130 ...  
## $ COMM\_Walk : Factor w/ 97 levels "1,200","1,215",..: 60 55 13 20 87 38 18 83 24 89 ...  
## $ COMM\_Bicycle : Factor w/ 66 levels "0","1,075","1,145",..: 53 37 18 26 22 44 22 11 41 1 ...  
## $ COMM\_Other : int 95 170 20 55 30 65 65 55 45 30 ...  
## $ COMM\_Tot : Factor w/ 132 levels "10,050","10,100",..: 29 24 53 58 50 122 128 53 99 60 ...  
## $ EDU\_Tot : Factor w/ 132 levels "10,335","10,420",..: 40 39 70 75 65 9 8 68 106 74 ...  
## $ EDU\_None : Factor w/ 123 levels "1,005","1,025",..: 50 78 120 115 109 27 4 84 84 56 ...  
## $ EDU\_HS\_Lower : int 7355 9080 2310 2530 2445 4895 3905 1725 2175 1095 ...  
## $ EDU\_High\_School : Factor w/ 122 levels "1,005","1,015",..: 97 99 17 34 33 81 75 15 38 118 ...  
## $ EDU\_College\_Trades : int 4705 3800 1480 1730 1420 2905 2880 1595 1875 1020 ...  
## $ EDU\_University : int 5600 4135 1570 1505 1060 3315 4325 1950 4250 3520 ...  
## $ DWEL\_SD\_House : Factor w/ 126 levels "0","1,010","1,050",..: 85 21 17 26 126 44 82 23 54 73 ...  
## $ DWEL\_Apart : Factor w/ 123 levels "0","1,015","1,035",..: 40 99 109 87 11 80 69 24 44 86 ...  
## $ DWEL\_Attached : Factor w/ 125 levels "1,020","1,045",..: 89 41 1 23 119 120 6 102 10 37 ...  
## $ DWEL\_Movable : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ INCHH\_Under\_25000 : int 1195 1490 410 670 445 1060 990 500 620 190 ...  
## $ INCHH\_25000\_50000 : int 2130 2875 815 850 865 2025 1895 925 1140 375 ...  
## $ INCHH\_50000\_99999 : int 3870 3810 1140 1295 1180 2815 2740 1370 1905 830 ...  
## $ INCHH\_Over\_100000 : int 3085 1700 935 1035 720 1875 2915 1345 2580 2440 ...  
## $ INCHH\_25000\_99999 : int 6000 6685 1955 2145 2045 4840 4635 2295 3045 1205 ...  
## $ INCIND\_Pop\_with\_amount: Factor w/ 132 levels "10,060","10,165",..: 83 80 110 113 103 49 58 114 23 119 ...  
## $ INCIND\_Population\_15up: Factor w/ 134 levels "10,000","10,105",..: 90 87 118 121 110 57 63 120 26 126 ...  
## $ P\_INCIND\_Total\_Avg : Factor w/ 137 levels "100,516","101,551",..: 32 14 42 52 37 63 80 86 2 137 ...  
## $ INCIND\_pop\_low\_inc : Factor w/ 127 levels "1,000","1,030",..: 86 117 17 22 25 84 50 5 11 115 ...  
## $ P\_INCIND\_pop\_low\_inc : int 14 22 15 16 18 20 11 11 8 6 ...  
## $ HH\_1\_Person : Factor w/ 122 levels "1,025","1,105",..: 35 23 101 122 104 32 57 8 33 100 ...  
## $ HH\_2\_Persons : Factor w/ 126 levels "1,010","1,015",..: 78 59 121 4 122 70 85 16 59 15 ...  
## $ HH\_3\_Persons : Factor w/ 113 levels "1,005","1,010",..: 60 59 84 89 86 31 36 90 3 95 ...  
## $ HH\_4\_Persons : Factor w/ 117 levels "1,000","1,005",..: 48 49 74 83 72 19 31 78 111 113 ...  
## $ HH\_5\_Persons : Factor w/ 99 levels "1,050","1,065",..: 30 32 82 61 70 5 93 46 67 54 ...  
## $ P\_HH\_Avg\_Size : int 3 3 3 3 3 3 3 2 2 3 ...  
## $ HH\_Tot : Factor w/ 133 levels "10,050","10,060",..: 6 130 34 46 32 110 116 51 87 47 ...  
## $ IMM\_Before\_1981 : Factor w/ 120 levels "1,025","1,035",..: 101 77 9 17 7 84 94 39 68 19 ...  
## $ IMM\_1981\_2000 : int 7860 7205 2075 1895 2010 3650 3035 1705 1860 1065 ...  
## $ IMM\_2001\_2016 : int 7585 11815 2210 1755 1620 4950 3570 1235 1845 950 ...  
## $ IMM\_Non\_Imm : Factor w/ 137 levels "10,080","10,110",..: 18 6 54 63 57 2 16 67 126 106 ...  
## $ IMM\_Non\_Perm\_res : Factor w/ 107 levels "1,005","1,010",..: 13 100 32 23 22 69 51 21 31 23 ...  
## $ IMM\_No : int 13655 11325 4660 5385 4720 10560 12115 5470 9250 7695 ...  
## $ IMM\_Yes : int 19235 21485 5460 4970 4730 11435 10025 4645 5950 3350 ...  
## $ IND\_Cat\_11 : int 20 40 0 25 0 10 10 0 20 20 ...  
## $ IND\_Cat\_21 : int 20 10 10 0 0 15 10 0 15 10 ...  
## $ IND\_Cat\_22 : int 45 15 20 20 0 35 50 10 55 20 ...  
## $ IND\_Cat\_23 : Factor w/ 102 levels "1,000","1,015",..: 3 97 54 52 33 87 92 57 71 47 ...  
## $ IND\_Cat\_31 : Factor w/ 109 levels "1,000","1,005",..: 40 36 98 90 78 108 109 59 85 60 ...  
## $ IND\_Cat\_41 : Factor w/ 83 levels "1,020","1,035",..: 78 74 26 29 24 50 62 21 47 37 ...  
## $ IND\_Cat\_44 : Factor w/ 109 levels "1,010","1,035",..: 41 32 78 80 67 4 11 64 94 63 ...  
## $ IND\_Cat\_48 : Factor w/ 93 levels "1,045","1,225",..: 7 3 61 59 56 90 82 37 52 19 ...  
## $ IND\_Cat\_51 : Factor w/ 93 levels "1,005","1,025",..: 53 43 11 91 10 29 17 12 19 13 ...  
## $ IND\_Cat\_52 : Factor w/ 111 levels "1,015","1,020",..: 91 61 33 36 33 57 99 47 96 67 ...  
## $ IND\_Cat\_53 : Factor w/ 73 levels "1,075","1,625",..: 37 17 68 3 72 17 29 10 40 34 ...  
## $ IND\_Cat\_54 : Factor w/ 117 levels "1,000","1,005",..: 115 84 61 67 59 92 108 73 8 113 ...  
## $ IND\_Cat\_55 : int 35 30 0 0 20 30 35 0 15 20 ...  
## $ IND\_Cat\_56 : Factor w/ 106 levels "1,010","1,060",..: 8 5 36 60 43 94 79 33 44 19 ...  
## $ IND\_Cat\_61 : Factor w/ 104 levels "1,040","1,055",..: 94 59 28 34 29 63 96 50 67 62 ...  
## $ IND\_Cat\_62 : Factor w/ 110 levels "1,005","1,015",..: 36 23 58 70 64 2 100 62 95 70 ...  
## $ IND\_Cat\_71 : Factor w/ 72 levels "1,470","100",..: 41 24 62 70 55 25 30 72 14 6 ...  
## $ IND\_Cat\_72 : Factor w/ 112 levels "1,085","1,095",..: 10 7 48 42 41 100 65 40 55 36 ...  
## $ IND\_Cat\_81 : Factor w/ 89 levels "1,010","1,015",..: 75 65 10 14 11 48 57 26 41 16 ...  
## $ IND\_Cat\_91 : Factor w/ 81 levels "1,080","1,105",..: 57 33 10 19 10 31 44 17 44 34 ...  
## $ IND\_Tot : Factor w/ 132 levels "10,115","10,365",..: 37 32 62 65 58 130 5 63 112 75 ...  
## [list output truncated]

#solution: doesn't work on my version  
#u <- getURL("http://vincentarelbundock.github.io/Rdatasets/csv/Ecdat/Computers.csv")   
#c\_prices <- read.csv(text = u)

1. Create a multiple linear regression model to predict infection rate.
2. Split the dataset to 70% of training and 30% of test sets.

#The floor function rounds a numeric input down to the next lower integer  
rn\_train <- sample(nrow(covid\_dat), floor(nrow(covid\_dat)\*0.7))   
train <- covid\_dat[rn\_train,]   
test <- covid\_dat[-rn\_train,]  
  
str(covid\_dat)

## 'data.frame': 137 obs. of 226 variables:  
## $ NEIGHBOURHOOD\_ID : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ neighbourhood\_name : Factor w/ 137 levels "Agincourt North",..: 123 80 116 100 39 63 129 54 37 98 ...  
## $ TOT\_HOSPITALIZED : int 128 127 54 27 24 71 55 41 45 17 ...  
## $ TOT\_IN\_ICU : int 18 24 6 4 1 16 8 4 7 2 ...  
## $ TOT\_INTUBATED : int 7 17 1 1 0 13 3 3 3 0 ...  
## $ ACTIVE\_CASES : int 111 87 30 19 24 62 43 17 17 12 ...  
## $ FATAL\_CASES : int 50 20 56 21 2 17 8 50 18 2 ...  
## $ RESOLVED\_CASES : int 2441 2773 831 532 539 1247 828 460 474 221 ...  
## $ TOTAL\_CASES : int 2602 2880 917 572 565 1326 879 527 509 235 ...  
## $ INFECT\_CLOSE\_CONT : int 152 186 59 42 40 90 60 23 28 13 ...  
## $ INFECT\_OB\_CONGR : int 8 15 25 2 8 17 1 5 1 3 ...  
## $ INFECT\_OB\_HEALTH : int 214 82 180 89 20 39 46 192 112 10 ...  
## $ INFECT\_OB\_OTHER : int 122 109 23 23 13 54 37 19 17 13 ...  
## $ INFECT\_NO\_INFO : int 865 1068 271 187 193 409 280 123 147 78 ...  
## $ INFECT\_COMMUNITY : int 693 773 190 133 161 412 285 103 127 63 ...  
## $ INFECT\_TRAVEL : int 30 20 4 4 7 6 4 3 6 6 ...  
## $ INFECT\_PENDING : int 1 4 0 3 0 3 4 3 0 3 ...  
## $ P\_FATALITY\_RATE : num 0.01922 0.00694 0.06107 0.03671 0.00354 ...  
## $ P\_Hospital\_RATE : num 0.0492 0.0441 0.0589 0.0472 0.0425 ...  
## $ P\_ICU\_RATE : num 0.00692 0.00833 0.00654 0.00699 0.00177 ...  
## $ P\_INFECT\_CLOSE\_CONT : num 0.0584 0.0646 0.0643 0.0734 0.0708 ...  
## $ P\_INFECT\_OB\_CONGR : num 0.00307 0.00521 0.02726 0.0035 0.01416 ...  
## $ P\_INFECT\_OB\_HEALTH : num 0.0822 0.0285 0.1963 0.1556 0.0354 ...  
## $ P\_INFECT\_OB\_OTHER : num 0.0469 0.0378 0.0251 0.0402 0.023 ...  
## $ P\_INFECT\_NO\_INFO : num 0.332 0.371 0.296 0.327 0.342 ...  
## $ P\_INFECT\_COMMUNITY : num 0.266 0.268 0.207 0.233 0.285 ...  
## $ P\_INFECT\_TRAVEL : num 0.01153 0.00694 0.00436 0.00699 0.01239 ...  
## $ P\_INFECT\_PENDING : num 0.000384 0.001389 0 0.005245 0 ...  
## $ NBH\_NIA\_IND : int 0 1 1 0 1 1 0 0 0 0 ...  
## $ NBH\_SH\_UNITS : int 950 1288 372 308 358 553 762 688 0 0 ...  
## $ NBH\_SH\_RGI : int 411 1181 180 299 358 401 511 390 0 0 ...  
## $ AGE1\_0\_to\_14 : Factor w/ 126 levels "1,120","1,150",..: 113 122 30 27 36 103 89 17 46 33 ...  
## $ AGE1\_15\_to\_24 : Factor w/ 128 levels "1,035","1,040",..: 113 112 20 16 24 85 79 7 43 31 ...  
## $ AGE1\_25\_to\_54 : Factor w/ 134 levels "10,310","10,350",..: 23 22 54 56 48 121 117 50 81 51 ...  
## $ AGE1\_55\_to\_64 : Factor w/ 122 levels "1,050","1,100",..: 106 93 5 21 9 71 80 20 64 44 ...  
## $ AGE1\_65\_to\_84 : Factor w/ 126 levels "1,025","1,095",..: 105 82 29 24 5 83 104 69 74 35 ...  
## $ AGE1\_85\_up : Factor w/ 93 levels "1,000","1,040",..: 68 34 41 34 12 65 86 92 73 39 ...  
## $ COMM\_Vehicle\_Driver : Factor w/ 128 levels "1,075","1,100",..: 127 119 49 53 34 101 115 47 92 71 ...  
## $ COMM\_Vehicle\_PASs : Factor w/ 82 levels "1,060","1,140",..: 2 80 39 24 27 55 62 21 42 23 ...  
## $ COMM\_Public\_Transit : Factor w/ 130 levels "1,030","1,085",..: 101 97 1 11 10 54 46 5 34 130 ...  
## $ COMM\_Walk : Factor w/ 97 levels "1,200","1,215",..: 60 55 13 20 87 38 18 83 24 89 ...  
## $ COMM\_Bicycle : Factor w/ 66 levels "0","1,075","1,145",..: 53 37 18 26 22 44 22 11 41 1 ...  
## $ COMM\_Other : int 95 170 20 55 30 65 65 55 45 30 ...  
## $ COMM\_Tot : Factor w/ 132 levels "10,050","10,100",..: 29 24 53 58 50 122 128 53 99 60 ...  
## $ EDU\_Tot : Factor w/ 132 levels "10,335","10,420",..: 40 39 70 75 65 9 8 68 106 74 ...  
## $ EDU\_None : Factor w/ 123 levels "1,005","1,025",..: 50 78 120 115 109 27 4 84 84 56 ...  
## $ EDU\_HS\_Lower : int 7355 9080 2310 2530 2445 4895 3905 1725 2175 1095 ...  
## $ EDU\_High\_School : Factor w/ 122 levels "1,005","1,015",..: 97 99 17 34 33 81 75 15 38 118 ...  
## $ EDU\_College\_Trades : int 4705 3800 1480 1730 1420 2905 2880 1595 1875 1020 ...  
## $ EDU\_University : int 5600 4135 1570 1505 1060 3315 4325 1950 4250 3520 ...  
## $ DWEL\_SD\_House : Factor w/ 126 levels "0","1,010","1,050",..: 85 21 17 26 126 44 82 23 54 73 ...  
## $ DWEL\_Apart : Factor w/ 123 levels "0","1,015","1,035",..: 40 99 109 87 11 80 69 24 44 86 ...  
## $ DWEL\_Attached : Factor w/ 125 levels "1,020","1,045",..: 89 41 1 23 119 120 6 102 10 37 ...  
## $ DWEL\_Movable : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ INCHH\_Under\_25000 : int 1195 1490 410 670 445 1060 990 500 620 190 ...  
## $ INCHH\_25000\_50000 : int 2130 2875 815 850 865 2025 1895 925 1140 375 ...  
## $ INCHH\_50000\_99999 : int 3870 3810 1140 1295 1180 2815 2740 1370 1905 830 ...  
## $ INCHH\_Over\_100000 : int 3085 1700 935 1035 720 1875 2915 1345 2580 2440 ...  
## $ INCHH\_25000\_99999 : int 6000 6685 1955 2145 2045 4840 4635 2295 3045 1205 ...  
## $ INCIND\_Pop\_with\_amount: Factor w/ 132 levels "10,060","10,165",..: 83 80 110 113 103 49 58 114 23 119 ...  
## $ INCIND\_Population\_15up: Factor w/ 134 levels "10,000","10,105",..: 90 87 118 121 110 57 63 120 26 126 ...  
## $ P\_INCIND\_Total\_Avg : Factor w/ 137 levels "100,516","101,551",..: 32 14 42 52 37 63 80 86 2 137 ...  
## $ INCIND\_pop\_low\_inc : Factor w/ 127 levels "1,000","1,030",..: 86 117 17 22 25 84 50 5 11 115 ...  
## $ P\_INCIND\_pop\_low\_inc : int 14 22 15 16 18 20 11 11 8 6 ...  
## $ HH\_1\_Person : Factor w/ 122 levels "1,025","1,105",..: 35 23 101 122 104 32 57 8 33 100 ...  
## $ HH\_2\_Persons : Factor w/ 126 levels "1,010","1,015",..: 78 59 121 4 122 70 85 16 59 15 ...  
## $ HH\_3\_Persons : Factor w/ 113 levels "1,005","1,010",..: 60 59 84 89 86 31 36 90 3 95 ...  
## $ HH\_4\_Persons : Factor w/ 117 levels "1,000","1,005",..: 48 49 74 83 72 19 31 78 111 113 ...  
## $ HH\_5\_Persons : Factor w/ 99 levels "1,050","1,065",..: 30 32 82 61 70 5 93 46 67 54 ...  
## $ P\_HH\_Avg\_Size : int 3 3 3 3 3 3 3 2 2 3 ...  
## $ HH\_Tot : Factor w/ 133 levels "10,050","10,060",..: 6 130 34 46 32 110 116 51 87 47 ...  
## $ IMM\_Before\_1981 : Factor w/ 120 levels "1,025","1,035",..: 101 77 9 17 7 84 94 39 68 19 ...  
## $ IMM\_1981\_2000 : int 7860 7205 2075 1895 2010 3650 3035 1705 1860 1065 ...  
## $ IMM\_2001\_2016 : int 7585 11815 2210 1755 1620 4950 3570 1235 1845 950 ...  
## $ IMM\_Non\_Imm : Factor w/ 137 levels "10,080","10,110",..: 18 6 54 63 57 2 16 67 126 106 ...  
## $ IMM\_Non\_Perm\_res : Factor w/ 107 levels "1,005","1,010",..: 13 100 32 23 22 69 51 21 31 23 ...  
## $ IMM\_No : int 13655 11325 4660 5385 4720 10560 12115 5470 9250 7695 ...  
## $ IMM\_Yes : int 19235 21485 5460 4970 4730 11435 10025 4645 5950 3350 ...  
## $ IND\_Cat\_11 : int 20 40 0 25 0 10 10 0 20 20 ...  
## $ IND\_Cat\_21 : int 20 10 10 0 0 15 10 0 15 10 ...  
## $ IND\_Cat\_22 : int 45 15 20 20 0 35 50 10 55 20 ...  
## $ IND\_Cat\_23 : Factor w/ 102 levels "1,000","1,015",..: 3 97 54 52 33 87 92 57 71 47 ...  
## $ IND\_Cat\_31 : Factor w/ 109 levels "1,000","1,005",..: 40 36 98 90 78 108 109 59 85 60 ...  
## $ IND\_Cat\_41 : Factor w/ 83 levels "1,020","1,035",..: 78 74 26 29 24 50 62 21 47 37 ...  
## $ IND\_Cat\_44 : Factor w/ 109 levels "1,010","1,035",..: 41 32 78 80 67 4 11 64 94 63 ...  
## $ IND\_Cat\_48 : Factor w/ 93 levels "1,045","1,225",..: 7 3 61 59 56 90 82 37 52 19 ...  
## $ IND\_Cat\_51 : Factor w/ 93 levels "1,005","1,025",..: 53 43 11 91 10 29 17 12 19 13 ...  
## $ IND\_Cat\_52 : Factor w/ 111 levels "1,015","1,020",..: 91 61 33 36 33 57 99 47 96 67 ...  
## $ IND\_Cat\_53 : Factor w/ 73 levels "1,075","1,625",..: 37 17 68 3 72 17 29 10 40 34 ...  
## $ IND\_Cat\_54 : Factor w/ 117 levels "1,000","1,005",..: 115 84 61 67 59 92 108 73 8 113 ...  
## $ IND\_Cat\_55 : int 35 30 0 0 20 30 35 0 15 20 ...  
## $ IND\_Cat\_56 : Factor w/ 106 levels "1,010","1,060",..: 8 5 36 60 43 94 79 33 44 19 ...  
## $ IND\_Cat\_61 : Factor w/ 104 levels "1,040","1,055",..: 94 59 28 34 29 63 96 50 67 62 ...  
## $ IND\_Cat\_62 : Factor w/ 110 levels "1,005","1,015",..: 36 23 58 70 64 2 100 62 95 70 ...  
## $ IND\_Cat\_71 : Factor w/ 72 levels "1,470","100",..: 41 24 62 70 55 25 30 72 14 6 ...  
## $ IND\_Cat\_72 : Factor w/ 112 levels "1,085","1,095",..: 10 7 48 42 41 100 65 40 55 36 ...  
## $ IND\_Cat\_81 : Factor w/ 89 levels "1,010","1,015",..: 75 65 10 14 11 48 57 26 41 16 ...  
## $ IND\_Cat\_91 : Factor w/ 81 levels "1,080","1,105",..: 57 33 10 19 10 31 44 17 44 34 ...  
## $ IND\_Tot : Factor w/ 132 levels "10,115","10,365",..: 37 32 62 65 58 130 5 63 112 75 ...  
## [list output truncated]

1. Create the multiple regression model on the train set using ram, screen, speed, hd and ads as independent variables.

We will train our model on the training set and do the prediction on the test set

model\_mlr<- lm(INFECTION\_RATE~P\_AGE\_40\_to\_64+P\_VM\_Yes+P\_EDU\_HS\_Lower+P\_HH\_5\_Persons+P\_OCC\_Ess\_Yes, data=train)  
model\_mlr

##   
## Call:  
## lm(formula = INFECTION\_RATE ~ P\_AGE\_40\_to\_64 + P\_VM\_Yes + P\_EDU\_HS\_Lower +   
## P\_HH\_5\_Persons + P\_OCC\_Ess\_Yes, data = train)  
##   
## Coefficients:  
## (Intercept) P\_AGE\_40\_to\_64 P\_VM\_Yes P\_EDU\_HS\_Lower P\_HH\_5\_Persons   
## 0.02891 -0.17124 -0.03190 -0.06490 0.12219   
## P\_OCC\_Ess\_Yes   
## 0.19406

1. Apply predict() function on the test set.

prediction <- predict(model\_mlr, interval="prediction", newdata =test)  
  
head(prediction)

## fit lwr upr  
## 5 0.06100956 0.04391028 0.07810885  
## 7 0.04482912 0.02784580 0.06181243  
## 11 0.03939335 0.02229282 0.05649388  
## 13 0.04843387 0.03128888 0.06557886  
## 22 0.07389919 0.05660883 0.09118956  
## 27 0.06386168 0.04662584 0.08109751

summary(prediction)

## fit lwr upr   
## Min. :0.00628 Min. :-0.011023 Min. :0.02358   
## 1st Qu.:0.02331 1st Qu.: 0.006218 1st Qu.:0.04041   
## Median :0.03982 Median : 0.022718 Median :0.05692   
## Mean :0.03908 Mean : 0.021805 Mean :0.05636   
## 3rd Qu.:0.05441 3rd Qu.: 0.037182 3rd Qu.:0.07159   
## Max. :0.07390 Max. : 0.056609 Max. :0.09119

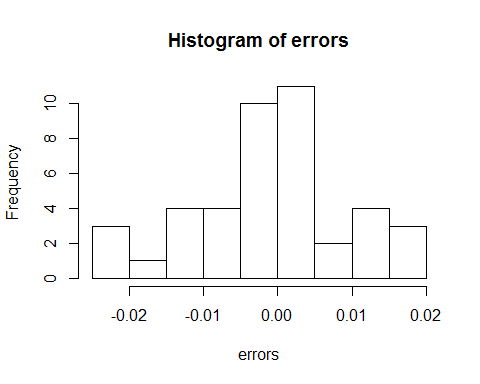
1. Calculate the errors (prediction price – test price) in predictions and show the histogram of errors.

Get difference between prediction and test

#this option for prediction "fit" model?   
#get the errors  
#this "fit" is a column, the fitted value/model. lower/upper are intervals upper/lower limit, we want to see model prediction. we want to compare the fitted value to the test  
summary(prediction)

## fit lwr upr   
## Min. :0.00628 Min. :-0.011023 Min. :0.02358   
## 1st Qu.:0.02331 1st Qu.: 0.006218 1st Qu.:0.04041   
## Median :0.03982 Median : 0.022718 Median :0.05692   
## Mean :0.03908 Mean : 0.021805 Mean :0.05636   
## 3rd Qu.:0.05441 3rd Qu.: 0.037182 3rd Qu.:0.07159   
## Max. :0.07390 Max. : 0.056609 Max. :0.09119

errors <- prediction[,"fit"] - test$INFECTION\_RATE   
hist(errors)



# The Error Histogram view can help you to understand how error is distributed  
#histogram is left scewed,   
  
  
errors2 <- prediction - test$price

1. Compute the root mean squared error.

need to get the error and then square it and get the summation and divide by number of observations in test set comparing the actual price to the predicted price

rmse <- sqrt(sum((prediction[,"fit"]-test$INFECTION\_RATE)^2)/nrow(test))   
rmse

## [1] 0.009956335

#both pred(25) and rmse values are better for multiple linear regression

##2. Apply forward selection and backward elimination algorithms to the dataset.

Forward selection:

we end up using all variables

#Forward SElection  
  
# for all variables  
  
  
full <- lm(INFECTION\_RATE~P\_AGE\_40\_to\_64+P\_VM\_Yes+P\_EDU\_HS\_Lower+P\_HH\_5\_Persons+P\_OCC\_Ess\_Yes,data=covid\_dat)   
  
# this creates for only one independent value at a time  
null <- lm(INFECTION\_RATE~1,data=covid\_dat)   
  
#stepAIC   
stepF <- stepAIC(null, scope=list(lower=null, upper=full), direction= "forward", trace=TRUE)

## Start: AIC=-1064.47  
## INFECTION\_RATE ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + P\_OCC\_Ess\_Yes 1 0.041384 0.015629 -1239.8  
## + P\_EDU\_HS\_Lower 1 0.031320 0.025693 -1171.7  
## + P\_HH\_5\_Persons 1 0.025502 0.031512 -1143.7  
## + P\_VM\_Yes 1 0.019046 0.037968 -1118.2  
## + P\_AGE\_40\_to\_64 1 0.001954 0.055060 -1067.2  
## <none> 0.057014 -1064.5  
##   
## Step: AIC=-1239.77  
## INFECTION\_RATE ~ P\_OCC\_Ess\_Yes  
##   
## Df Sum of Sq RSS AIC  
## + P\_EDU\_HS\_Lower 1 0.00151469 0.014115 -1251.7  
## + P\_AGE\_40\_to\_64 1 0.00147548 0.014154 -1251.3  
## + P\_HH\_5\_Persons 1 0.00040283 0.015227 -1241.3  
## <none> 0.015630 -1239.8  
## + P\_VM\_Yes 1 0.00004154 0.015588 -1238.1  
##   
## Step: AIC=-1251.73  
## INFECTION\_RATE ~ P\_OCC\_Ess\_Yes + P\_EDU\_HS\_Lower  
##   
## Df Sum of Sq RSS AIC  
## + P\_AGE\_40\_to\_64 1 0.00105879 0.013056 -1260.4  
## + P\_VM\_Yes 1 0.00039450 0.013720 -1253.6  
## + P\_HH\_5\_Persons 1 0.00029685 0.013818 -1252.6  
## <none> 0.014115 -1251.7  
##   
## Step: AIC=-1260.41  
## INFECTION\_RATE ~ P\_OCC\_Ess\_Yes + P\_EDU\_HS\_Lower + P\_AGE\_40\_to\_64  
##   
## Df Sum of Sq RSS AIC  
## + P\_VM\_Yes 1 0.00107499 0.011981 -1270.2  
## + P\_HH\_5\_Persons 1 0.00050936 0.012547 -1263.9  
## <none> 0.013056 -1260.4  
##   
## Step: AIC=-1270.18  
## INFECTION\_RATE ~ P\_OCC\_Ess\_Yes + P\_EDU\_HS\_Lower + P\_AGE\_40\_to\_64 +   
## P\_VM\_Yes  
##   
## Df Sum of Sq RSS AIC  
## + P\_HH\_5\_Persons 1 0.0017155 0.010266 -1289.4  
## <none> 0.011981 -1270.2  
##   
## Step: AIC=-1289.36  
## INFECTION\_RATE ~ P\_OCC\_Ess\_Yes + P\_EDU\_HS\_Lower + P\_AGE\_40\_to\_64 +   
## P\_VM\_Yes + P\_HH\_5\_Persons

summary(stepF)

##   
## Call:  
## lm(formula = INFECTION\_RATE ~ P\_OCC\_Ess\_Yes + P\_EDU\_HS\_Lower +   
## P\_AGE\_40\_to\_64 + P\_VM\_Yes + P\_HH\_5\_Persons, data = covid\_dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0177057 -0.0068203 0.0003322 0.0052340 0.0224149   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.015289 0.009263 1.651 0.101   
## P\_OCC\_Ess\_Yes 0.221844 0.020689 10.723 < 2e-16 \*\*\*  
## P\_EDU\_HS\_Lower -0.085779 0.017925 -4.786 4.53e-06 \*\*\*  
## P\_AGE\_40\_to\_64 -0.145097 0.024396 -5.948 2.33e-08 \*\*\*  
## P\_VM\_Yes -0.032584 0.006039 -5.395 3.10e-07 \*\*\*  
## P\_HH\_5\_Persons 0.112923 0.024135 4.679 7.10e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.008852 on 131 degrees of freedom  
## Multiple R-squared: 0.8199, Adjusted R-squared: 0.8131   
## F-statistic: 119.3 on 5 and 131 DF, p-value: < 2.2e-16

1. Apply backward elimination algorithms to the dataset.

#backward elimination  
  
full <- lm(INFECTION\_RATE~P\_AGE\_40\_to\_64+P\_VM\_Yes+P\_EDU\_HS\_Lower+P\_HH\_5\_Persons+P\_OCC\_Ess\_Yes,data=covid\_dat)   
  
stepB <- stepAIC(full, direction= "backward", trace=TRUE)

## Start: AIC=-1289.36  
## INFECTION\_RATE ~ P\_AGE\_40\_to\_64 + P\_VM\_Yes + P\_EDU\_HS\_Lower +   
## P\_HH\_5\_Persons + P\_OCC\_Ess\_Yes  
##   
## Df Sum of Sq RSS AIC  
## <none> 0.010266 -1289.4  
## - P\_HH\_5\_Persons 1 0.0017155 0.011981 -1270.2  
## - P\_EDU\_HS\_Lower 1 0.0017946 0.012060 -1269.3  
## - P\_VM\_Yes 1 0.0022811 0.012547 -1263.9  
## - P\_AGE\_40\_to\_64 1 0.0027720 0.013038 -1258.6  
## - P\_OCC\_Ess\_Yes 1 0.0090103 0.019276 -1205.0

summary(stepB)

##   
## Call:  
## lm(formula = INFECTION\_RATE ~ P\_AGE\_40\_to\_64 + P\_VM\_Yes + P\_EDU\_HS\_Lower +   
## P\_HH\_5\_Persons + P\_OCC\_Ess\_Yes, data = covid\_dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0177057 -0.0068203 0.0003322 0.0052340 0.0224149   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.015289 0.009263 1.651 0.101   
## P\_AGE\_40\_to\_64 -0.145097 0.024396 -5.948 2.33e-08 \*\*\*  
## P\_VM\_Yes -0.032584 0.006039 -5.395 3.10e-07 \*\*\*  
## P\_EDU\_HS\_Lower -0.085779 0.017925 -4.786 4.53e-06 \*\*\*  
## P\_HH\_5\_Persons 0.112923 0.024135 4.679 7.10e-06 \*\*\*  
## P\_OCC\_Ess\_Yes 0.221844 0.020689 10.723 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.008852 on 131 degrees of freedom  
## Multiple R-squared: 0.8199, Adjusted R-squared: 0.8131   
## F-statistic: 119.3 on 5 and 131 DF, p-value: < 2.2e-16

1. Apply regsubsets() to see the best combination of the attributes.

In the output, \* denotes the variables included to the subset. For example, the best combination of 4 attributes is: ram, speed, screen and trend, and the best combination of 5 attributes is: ram, hd, speed, screen and trend

This algorithm uses AIC at the back. putting all the combinatons and seeing which one has lower AIC to be better.  
part of a package.

# specify the model that is being used  
subsets<-regsubsets(INFECTION\_RATE~P\_AGE\_40\_to\_64+P\_VM\_Yes+P\_EDU\_HS\_Lower+P\_HH\_5\_Persons+P\_OCC\_Ess\_Yes, data=covid\_dat, nbest=1)   
  
sub.sum <- summary(subsets)   
  
as.data.frame(sub.sum$outmat)

## P\_AGE\_40\_to\_64 P\_VM\_Yes P\_EDU\_HS\_Lower P\_HH\_5\_Persons P\_OCC\_Ess\_Yes  
## 1 ( 1 ) \*  
## 2 ( 1 ) \* \*  
## 3 ( 1 ) \* \* \*  
## 4 ( 1 ) \* \* \* \*  
## 5 ( 1 ) \* \* \* \* \*

Random Forest

library(rsample)  
library(randomForest)

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

#library(ranger)  
#library(caret)  
#library(h2o)  
  
#set random seed to make results reproducible  
  
set.seed(17)  
  
#calculate the size of each of tre data sets;  
data\_set\_size <- floor(nrow(covid\_dat)/2)  
  
#generate a random sample of "data\_set\_size" indexes  
indexes <- sample(1:nrow(covid\_dat), size = data\_set\_size)  
  
#assign the data to the correct sets  
training = covid\_dat[indexes,]  
validation1 = covid\_dat[-indexes,]  
  
  
# Create training (70%) and test (30%) sets for the AmesHousing::make\_ames() data.  
# Use set.seed for reproducibility  
  
set.seed(123)  
data\_split <- initial\_split(covid\_dat, prop = .7)  
data\_train <- training(data\_split)  
data\_test <- testing(data\_split)

#import the package  
library(randomForest)  
  
class(covid\_dat$INFECTION\_RATE2)

## [1] "NULL"

#perform training'  
  
INFECTION\_RATE3<-factor(data\_train$INFECTION\_RATE2)  
class(INFECTION\_RATE3)

## [1] "factor"

rf\_infect\_rate <- randomForest(INFECTION\_RATE ~P\_AGE\_40\_to\_64+P\_VM\_Yes+P\_EDU\_HS\_Lower+P\_HH\_5\_Persons+P\_OCC\_Ess\_Yes, data=train, ntree=27)  
  
print(rf\_infect\_rate)

##   
## Call:  
## randomForest(formula = INFECTION\_RATE ~ P\_AGE\_40\_to\_64 + P\_VM\_Yes + P\_EDU\_HS\_Lower + P\_HH\_5\_Persons + P\_OCC\_Ess\_Yes, data = train, ntree = 27)   
## Type of random forest: regression  
## Number of trees: 27  
## No. of variables tried at each split: 1  
##   
## Mean of squared residuals: 0.0001105144  
## % Var explained: 72.35

#Run on testing  
library(MASS)  
library(ModelMetrics)

##   
## Attaching package: 'ModelMetrics'

## The following object is masked from 'package:base':  
##   
## kappa

res = lapply(c(111,222),function(i){  
 set.seed(i)  
  
fit = randomForest(INFECTION\_RATE ~P\_AGE\_40\_to\_64+P\_VM\_Yes+P\_EDU\_HS\_Lower+P\_HH\_5\_Persons+P\_OCC\_Ess\_Yes, data=train, ntree=27)  
   
pred\_values = predict(fit,test)  
actual\_values = test$INFECTION\_RATE  
  
#rmse(pred\_values,actual\_values)  
  
data.frame(seed=i,  
 metrics\_rmse = rmse(pred\_values,actual\_values),  
 cal\_rmse=mean((pred\_values-actual\_values)^2)^0.5  
 )  
})  
res=do.call(rbind,res)  
head(res)

## seed metrics\_rmse cal\_rmse  
## 1 111 0.01156428 0.01156428  
## 2 222 0.01265841 0.01265841

# Output to be present as PNG file   
png(file = "Q:/BI Team/Credit Card Reporting/Projects/ADH0820/Charts/randomForestRegression.png")  
   
# Plot the error vs the number of trees graph  
plot(rf\_infect\_rate)  
  
  
# Saving the file  
dev.off()

## png   
## 2

varImpPlot(rf\_infect\_rate)

