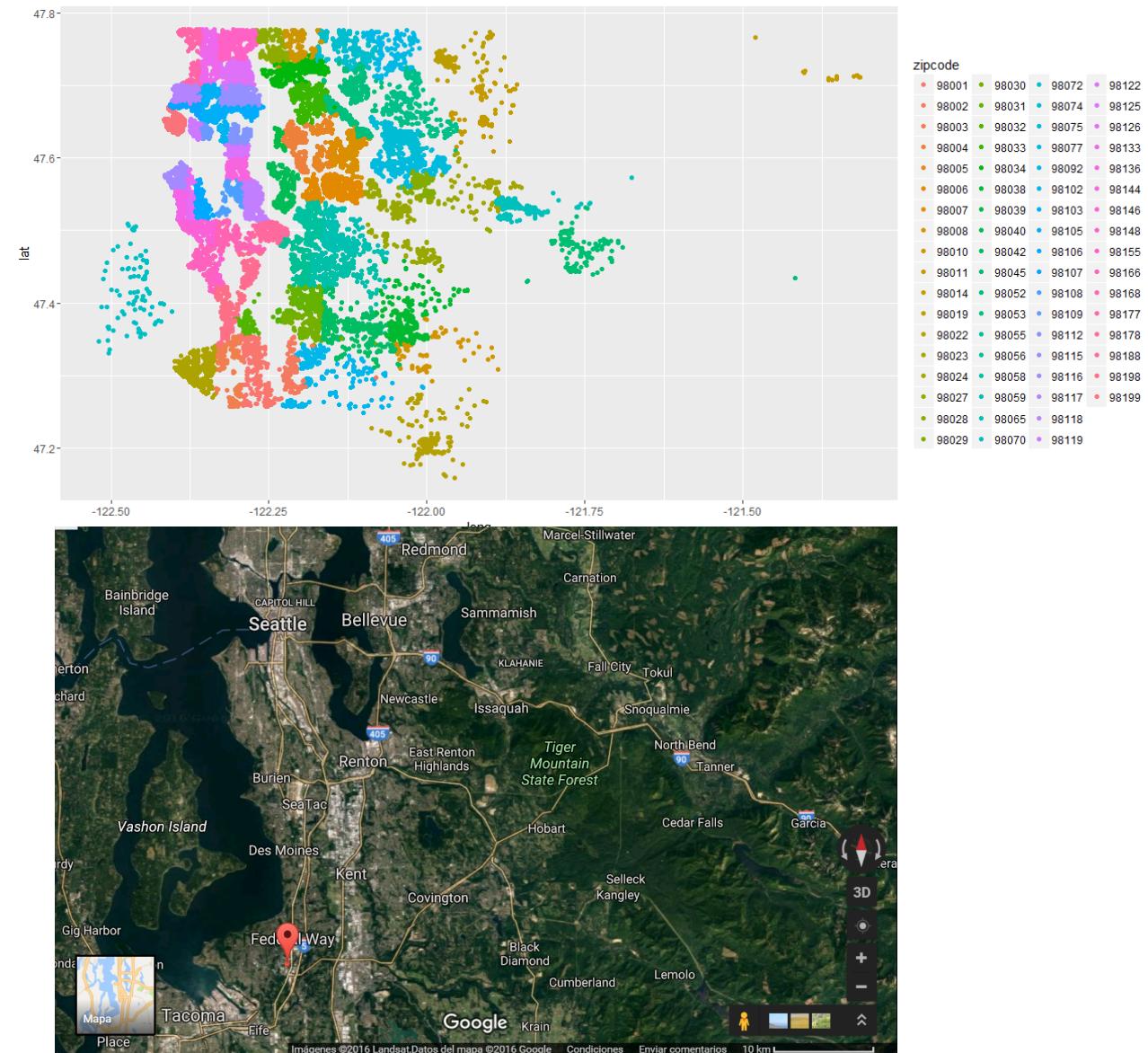


Primer Ejercicio: Construcción de un modelo de regresión.

Análisis previo

En primer lugar analizaré los datos de las viviendas del data set.

Graficando la ubicación y el código postal de las viviendas podemos observar que todas se encuentran en Seattle.



Voy a analizar otras características de las viviendas, empezando por los campos conocidos. Adicionalmente averiguaré el significado de las demás variables para poder evaluar su utilidad en el modelo.

Las variables conocidas son:

Id: identificador de la vivienda

date: fecha asociada a la información

price: precio de la vivienda

bedrooms: número de habitaciones

bathrooms: número de baños
sqft_living=superficie de la vivienda (en pies)
sqft_lot: superficie de la parcela (en pies)
floors: número de plantas
waterfront: indicador de estancia en primera línea al mar
view: número de orientaciones de la vivienda
yr_built: año de construcción
yr_renovated: año de reforma
zipcode: código postal
lat: latitud
long: longitud

Busco el significado de las variables desconocidas:

condition:

Relative to age and grade. Coded 1-5.

1 = Poor- Worn out. Repair and overhaul needed on painted surfaces, roofing, plumbing, heating and numerous functional inadequacies. Excessive deferred maintenance and abuse, limited value-in-use, approaching abandonment or major reconstruction; reuse or change in occupancy is imminent. Effective age is near the end of the scale regardless of the actual chronological age.

2 = Fair- Badly worn. Much repair needed. Many items need refinishing or overhauling, deferred maintenance obvious, inadequate building utility and systems all shortening the life expectancy and increasing the effective age.

3 = Average- Some evidence of deferred maintenance and normal obsolescence with age in that a few minor repairs are needed, along with some refinishing. All major components still functional and contributing toward an extended life expectancy. Effective age and utility is standard for like properties of its class and usage.

4 = Good- No obvious maintenance required but neither is everything new. Appearance and utility are above the standard and the overall effective age will be lower than the typical property.

5= Very Good- All items well maintained, many having been overhauled and repaired as they have shown signs of wear, increasing the life expectancy and lowering the effective age with little deterioration or obsolescence evident with a high degree of utility.

grade:

Represents the construction quality of improvements. Grades run from grade 1 to 13. Generally defined as:

1-3 Falls short of minimum building standards. Normally cabin or inferior structure.

4 Generally older, low quality construction. Does not meet code.

5 Low construction costs and workmanship. Small, simple design.

6 Lowest grade currently meeting building code. Low quality materials and simple designs.

7 Average grade of construction and design. Commonly seen in plats and older sub-divisions.

8 Just above average in construction and design. Usually better materials in both the exterior and interior finish work.

9 Better architectural design with extra interior and exterior design and quality.

10 Homes of this quality generally have high quality features. Finish work is better and more design quality is seen in the floor plans. Generally have a larger square footage.

11 Custom design and higher quality finish work with added amenities of solid woods, bathroom fixtures and more luxurious options.

12 Custom design and excellent builders. All materials are of the highest quality and all conveniences are present.

13 Generally custom designed and built. Mansion level. Large amount of highest quality cabinet work, wood trim, marble, entry ways etc.

sqft_above: = sqft_living - sqft_basement

sqft_basement:

sqft_living15: the average house square footage of the 15 closest neighbours

sqft_lot15: the average lot square footage of the 15 closest neighbours

Fuente:

<http://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r>

<https://www.kaggle.com/harlfoxem/d/harlfoxem/housesalesprediction/>

Voy a formatear el data set, para que las variables categóricas se representen como factores.

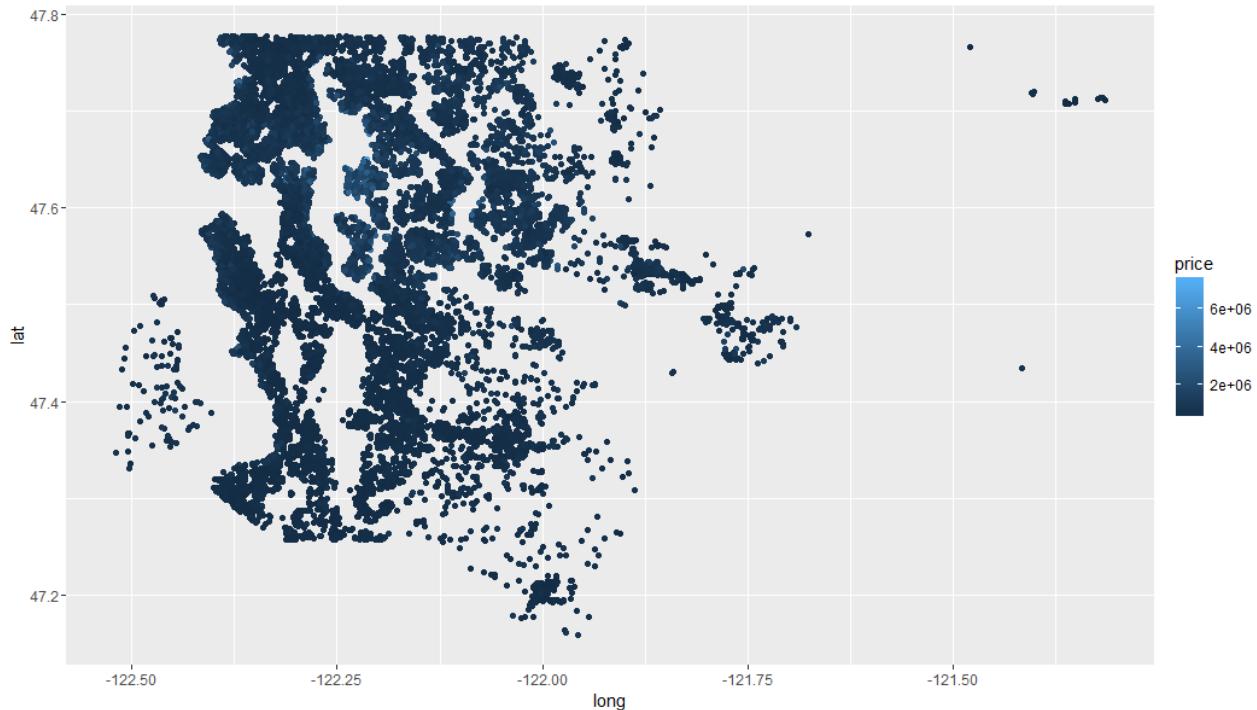
Tomo las siguientes variables como categóricas: bedrooms, bathrooms, floors, waterfront, view, condition, grade

El resto de las variables interesantes las dejo en formato numérico:

yr_built, yr_renovated, price, sqft_living , sqft_lot.

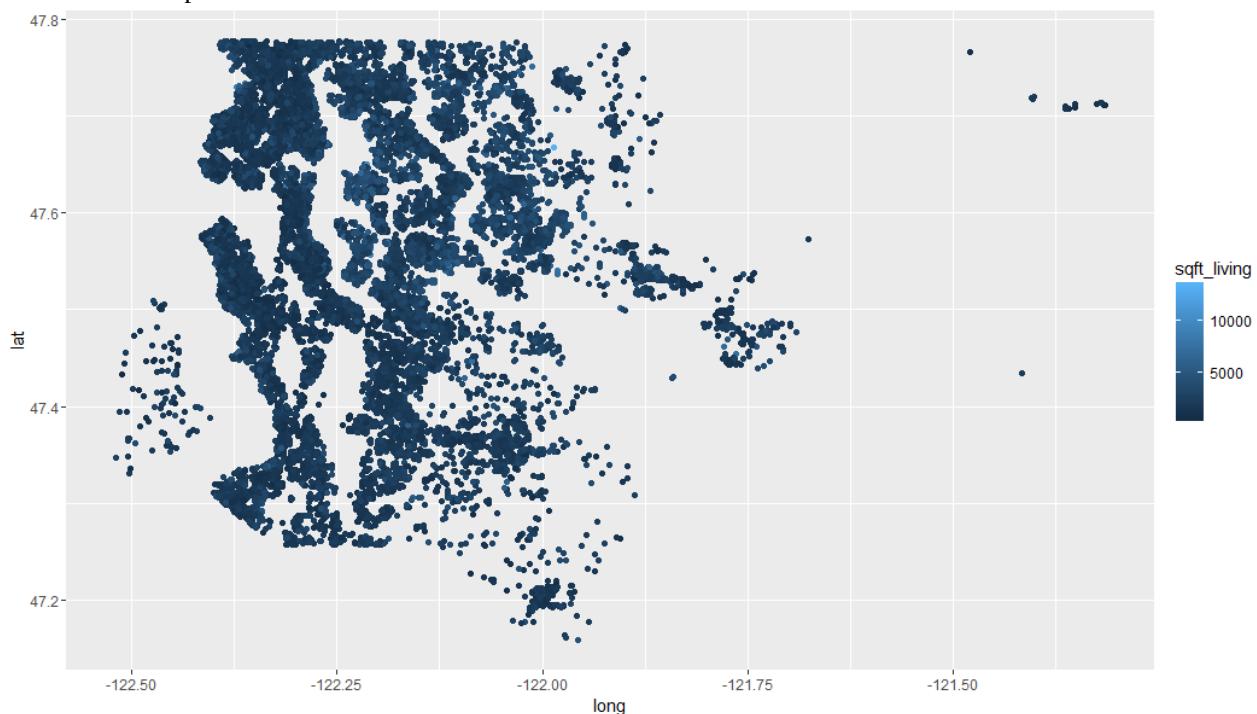
Análisis del efecto de la superficie de la vivienda en el precio de la vivienda

Veamos los precios de las viviendas:



El precio no parece tener mucha dispersión por zona, excepto el centro en alrededores de Capitol Hill y barrio Bellevue. Ahí las casas son mucho mas caras.

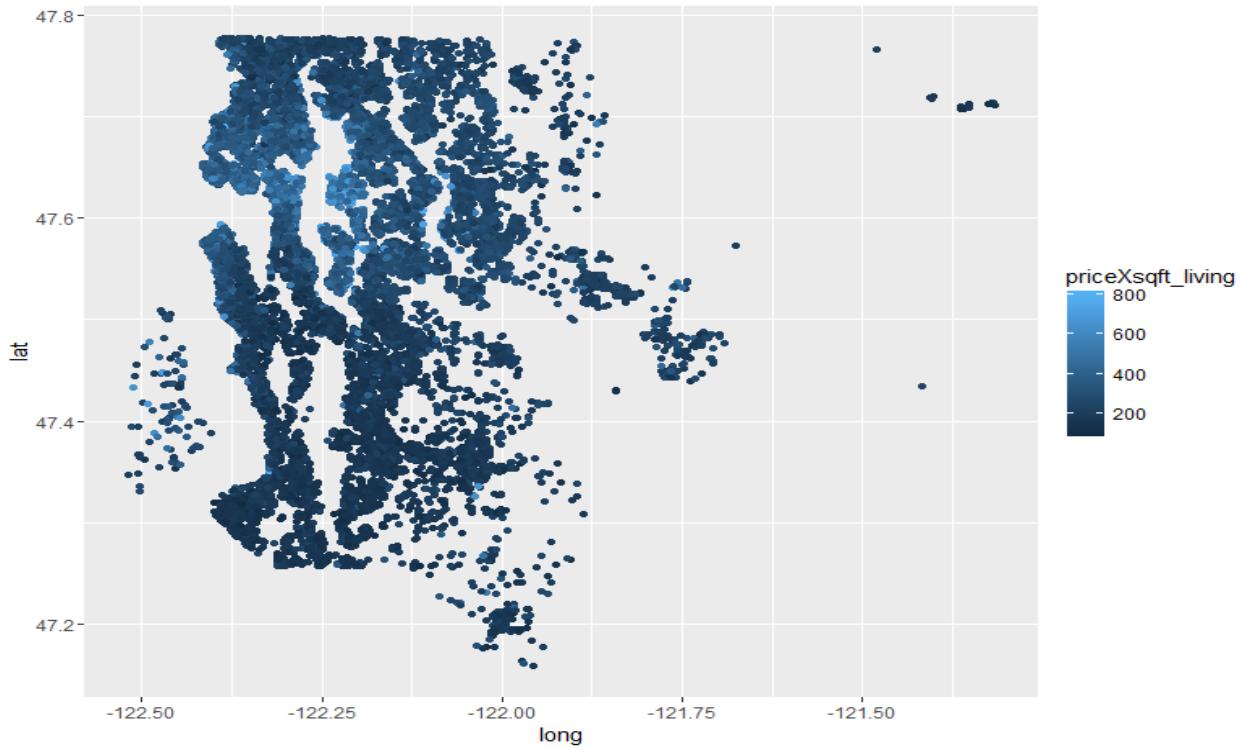
Analizando la superficie:



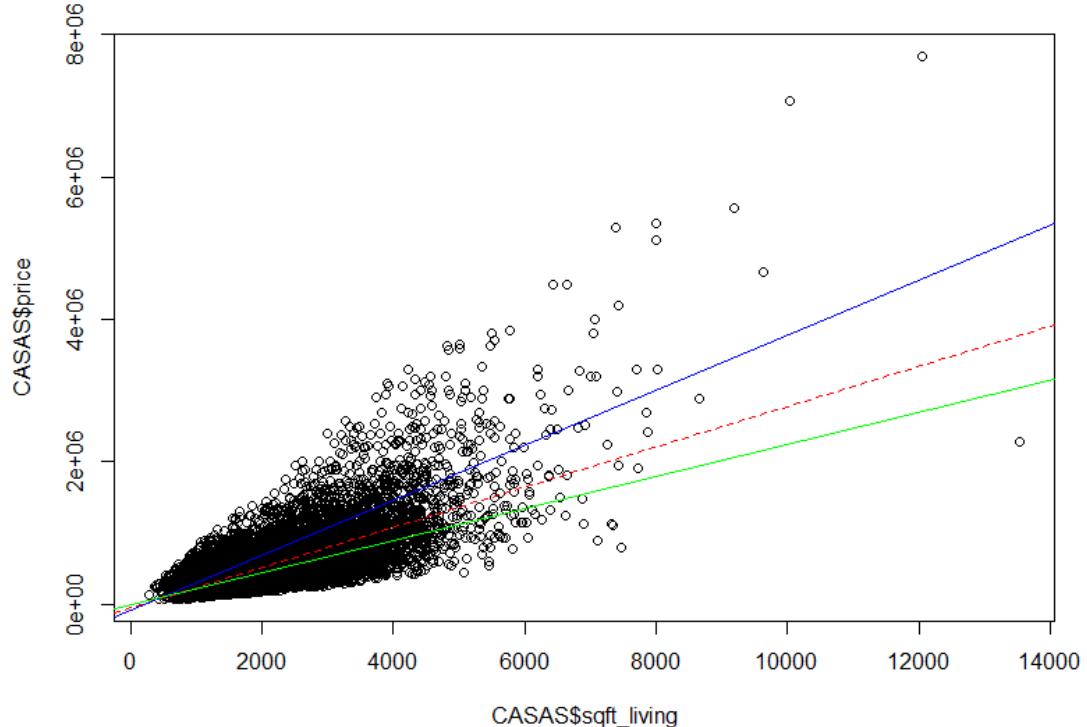
Se puede ver que en la zona cara las casas son más grandes. En general las casas son más grandes en el este y más pequeñas en el oeste.

Podemos observar que el precio no tiene relación directamente con la superficie. En algunas zonas las casas son pequeñas pero caras. Con toda seguridad la superficie no es el único factor del que depende el precio de las viviendas.

Veamos el precio por metro cuadrado (más bien por pie al cuadrado):



Es fácil de ver que hay una zona más cara que las demás. En alrededores del centro cada metro de superficie cuesta mucho más que en las demás zonas. Esta métrica indica bastante sobre la relación sobre el precio y la superficie marginal que queremos obtener. Es posible que existan dos patrones diferentes. Voy a comprobarlo:



El gráfico representa la relación entre el precio y superficie de la vivienda:

en rojo – para todas las viviendas

en azul – viviendas en el centro

en verde – viviendas en afueras

Podemos observar que la pendiente que determina la relación entre la superficie y el precio cambia dependiendo de la ubicación de la vivienda. En el caso de las viviendas del centro cada metro cuadrado adicional es mas caro que en el caso de las viviendas que están en afueras.

Adicionalmente realizaré un chow test para confirmar el cambio de la estructura:

```
chow.test(CASAS_CENTRO$price, CASAS_CENTRO$sqft_living, CASAS_AFUERAS$price, CASAS_AFUERAS$sqft_living)
```

F value	d.f.1	d.f.2	P value
3827.057	2.000	17343.000	0.000

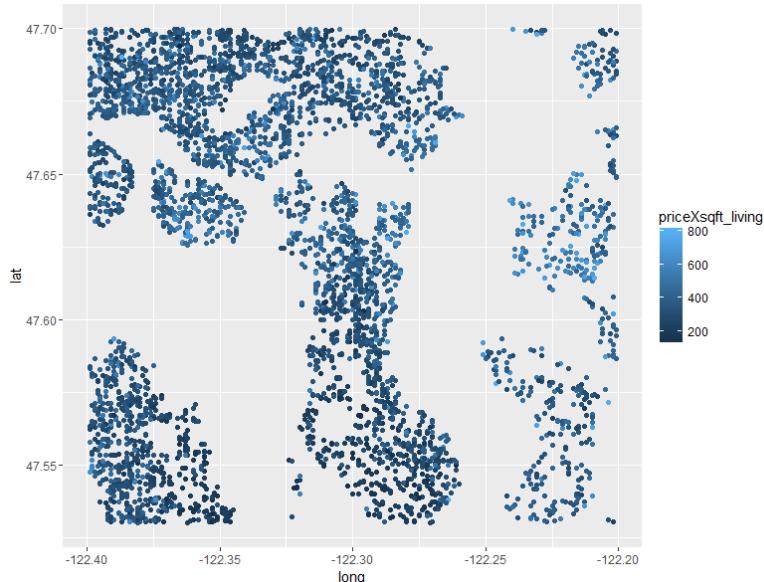
```
chow.test(log(CASAS_CENTRO$price), CASAS_CENTRO$sqft_living, log(CASAS_AFUERAS$price), CASAS_AFUERAS$sqft_living)
```

F value	d.f.1	d.f.2	P value
4033.446	2.000	17343.000	0.000

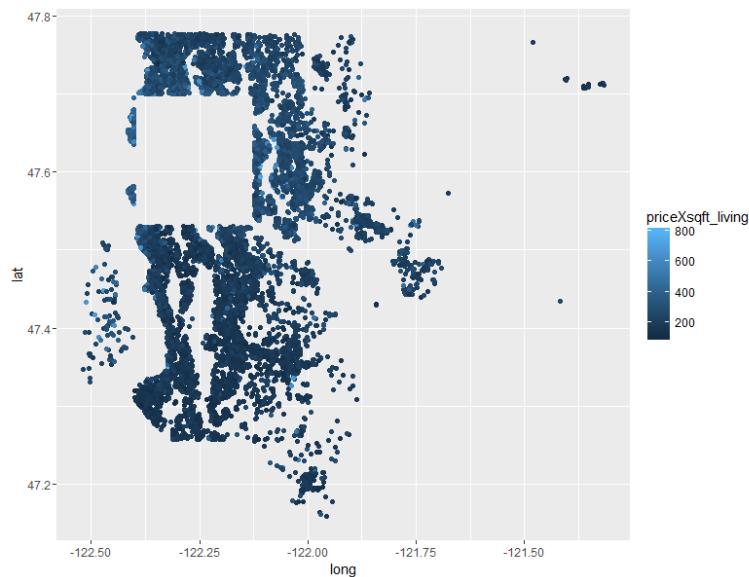
Se confirma la diferencia estructural y concluyo que los dos data sets probablemente tengan elasticidad del precio a la variación de superficie diferente.

Voy a dividir la población en dos partes:

Viviendas en el centro:



Viviendas en afuera:



Construcción del modelo básico para las viviendas del centro y para las viviendas en afuera.

Como el objetivo del análisis es construir un modelo que refleje el efecto de superficie en el precio de la vivienda voy a construir un modelo log. De esta forma podré estimar el impacto marginal de cada pie adicional de superficie a la variación del precio en %.

Para las casas del centro construyo el siguiente modelo básico:

```
modelocentrolog=lm(log(price)~sqft_living,data=CASAS_CENTRO)
```

```
lm(formula = log(price) ~ sqft_living, data = CASAS_CENTRO)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.44815	-0.16990	0.00652	0.17990	0.96124

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.253e+01	8.953e-03	1399.2	<2e-16 ***
sqft_living	3.965e-04	3.944e-06	100.5	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2933 on 5663 degrees of freedom

Intervalo de confianza:

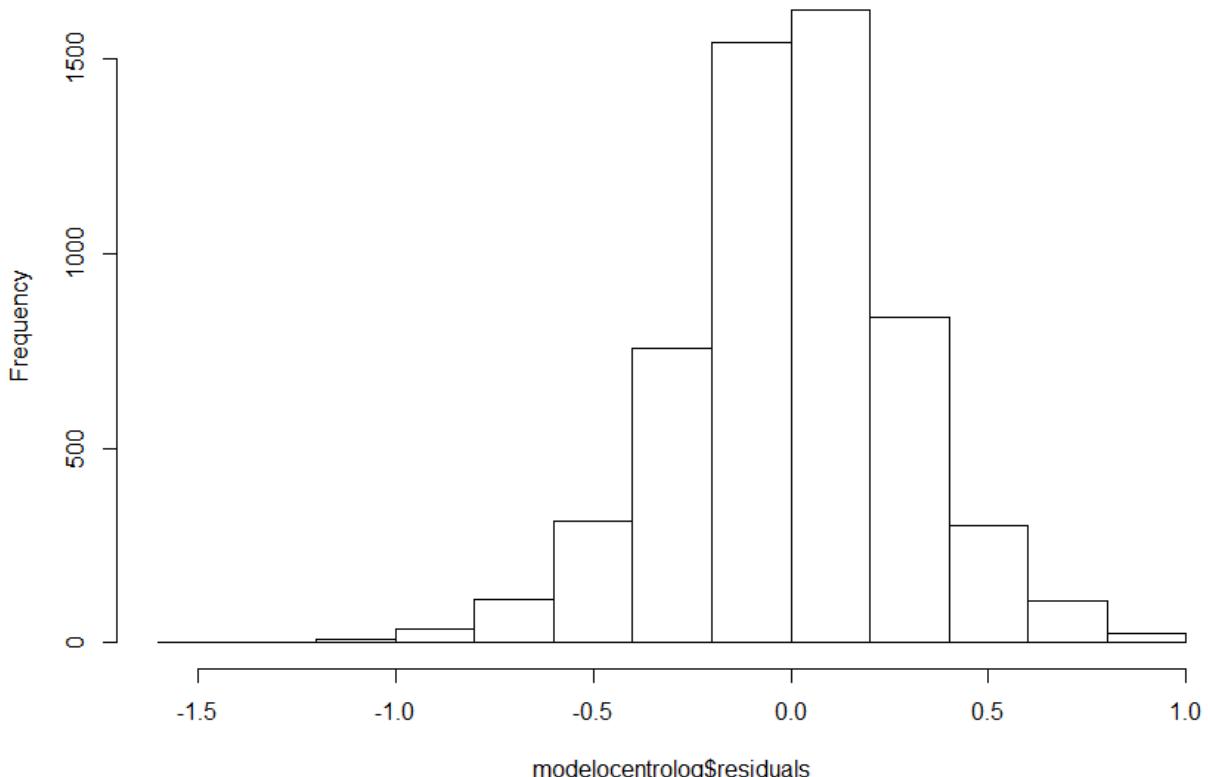
	2.5 %	97.5 %
(Intercept)	1.250929e+01	1.254439e+01
sqft_living	3.887853e-04	4.042502e-04

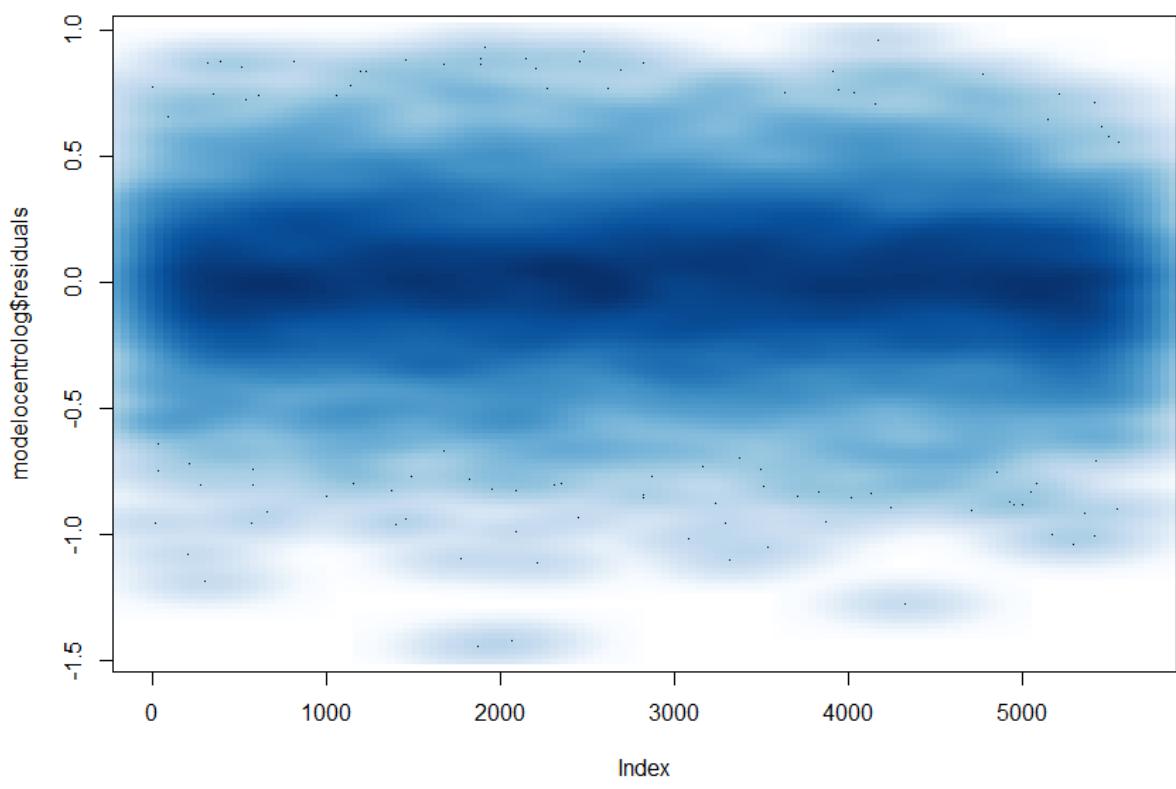
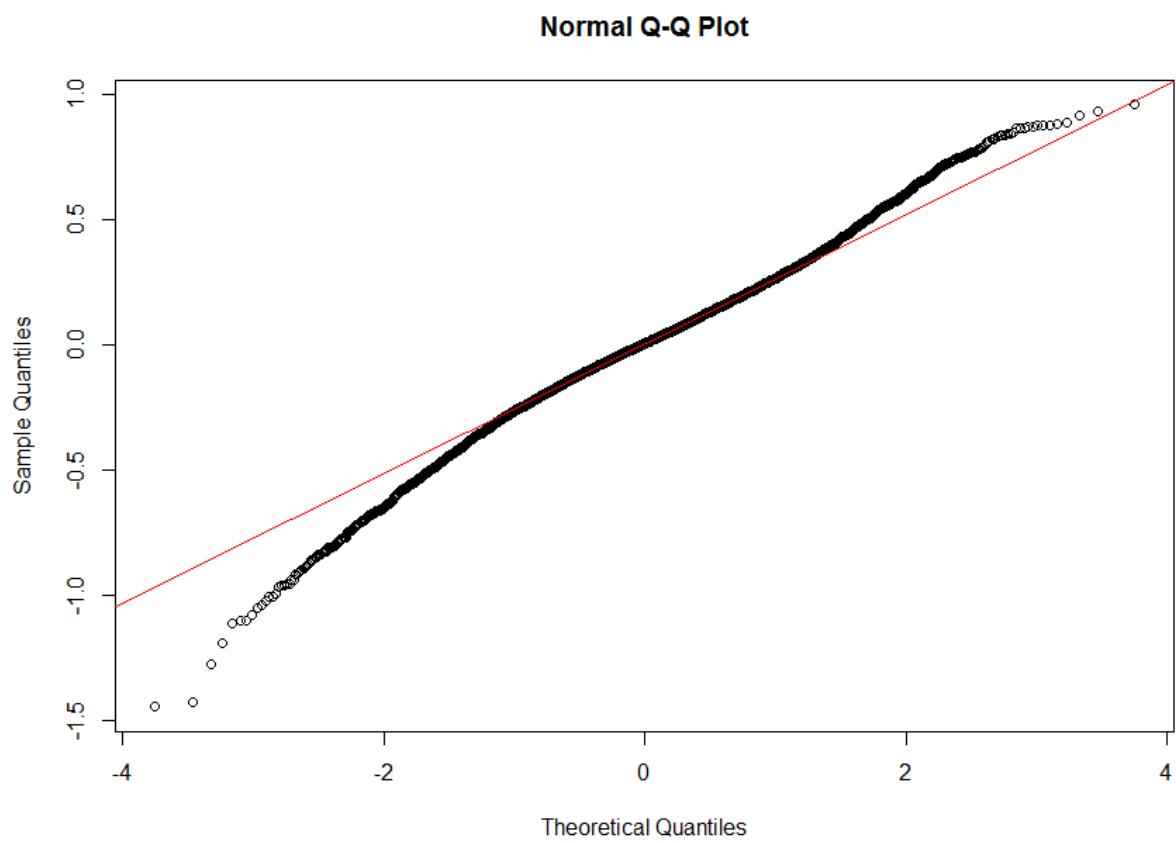
Es decir: ascenso de la superficie de las casas en el centro por 1 pie cuadrado está asociado con un ascenso en el precio entre 12,50388% (12,5%+ 0,00388%) y 12,54425% (12,54+0,00425%) con nivel de confianza de 95%.

El modelo tiene error residual estándar de 0.2933.

Residuos tienen media cercana a cero y distribución aproximada a normal en el segundo y tercer quartil teórico, aunque observo una desviación en las colas:

Histogram of modelocentrolog\$residuals





Voy a comprobar si el desajuste del modelo podría ser por outliers, construyendo un modelo robusto para comparar.

```
Call: rlm(formula = log(price) ~ sqft_living, data = CASAS_CENTRO)
Residuals:
    Min      1Q  Median      3Q     Max 
-1.468688 -0.173727  0.002251  0.176189  0.956280 

Coefficients:
            value   std. Error t value
(Intercept) 12.5277   0.0085 1468.6994
sqft_living  0.0004   0.0000 105.9491

Residual standard error: 0.2592 on 5663 degrees of freedom
```

Intervalo de confianza:

	2.5 %	97.5 %
(Intercept)	1.251098e+01	1.254442e+01
sqft_living	3.907854e-04	4.055163e-04

El error del modelo robusto baja pero no es una bajada considerable. El intervalo de confianza también es mas estrecho, pero no mucho más.

Comparando la bondad de ambos modelos obtengo:

	modelo centro log	modelo centro log robusto
AIC	2182.275	2183.608
BIC	2202.201	2203.535

Concluyo que el modelo no mejora mediante tratamiento de outliers. Más adelante voy a ver si se puede mejorar añadiendo más variables.

Para las casas de las afueras construyo el siguiente modelo básico:

```
modeloafuera=lm(log(price)~sqft_living,data=CASAS_AFUERAS)
```

```
Call:
lm(formula = log(price) ~ sqft_living, data = CASAS_AFUERAS)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.93030 -0.23043 -0.00428  0.22854  1.41924 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.205e+01  7.653e-03 1574.4   <2e-16 ***
sqft_living 4.078e-04  3.360e-06 121.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3222 on 11680 degrees of freedom
```

Intervalo de confianza:

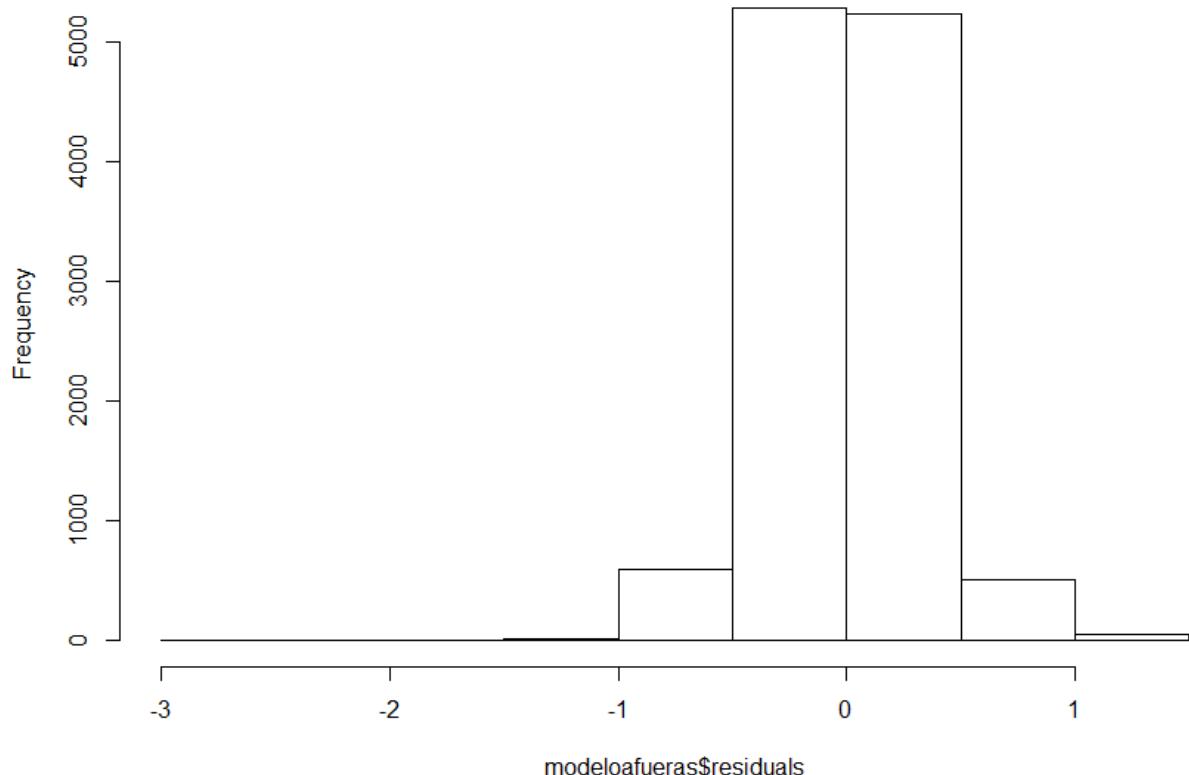
	2.5 %	97.5 %
(Intercept)	1.203333e+01	1.206333e+01
sqft_living	4.012175e-04	4.143893e-04

Es decir: ascenso de la superficie de las casas en el centro por 1 pie cuadrado está asociado con un ascenso en el precio entre 12,037% (12,033%+ 0,00401%) y 12,094% (12,063+0,00414%) con nivel de confianza de 95%.

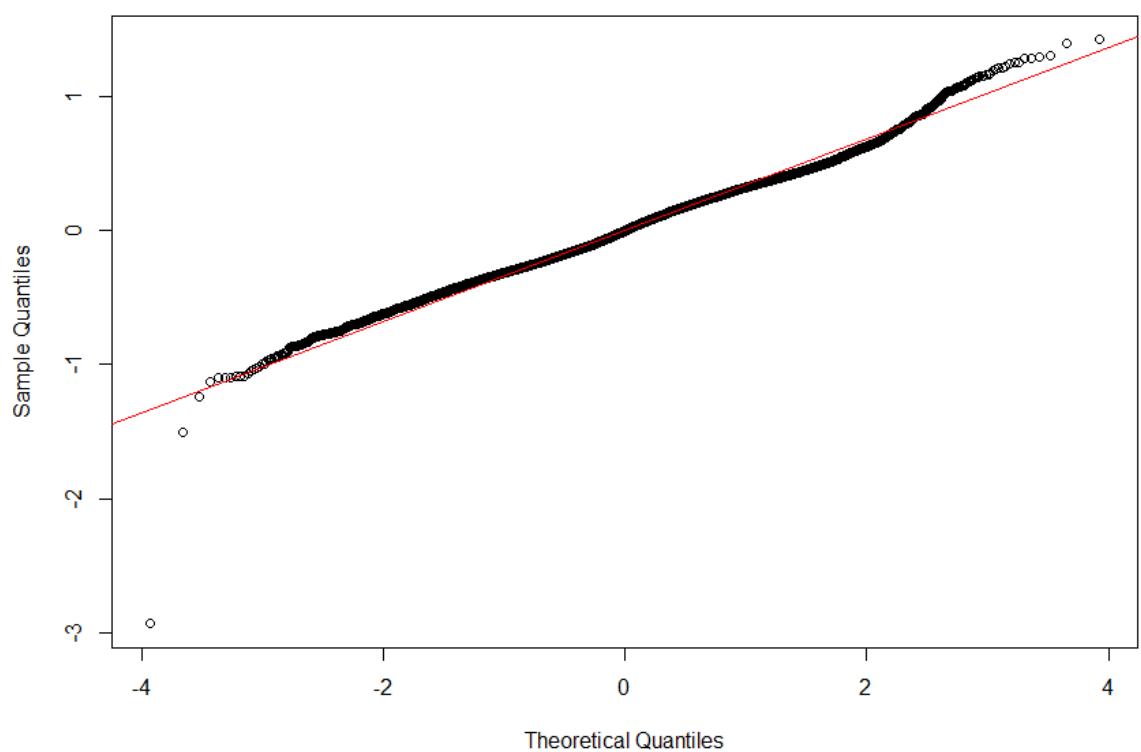
El modelo tiene error residual estándar de 0.3222.

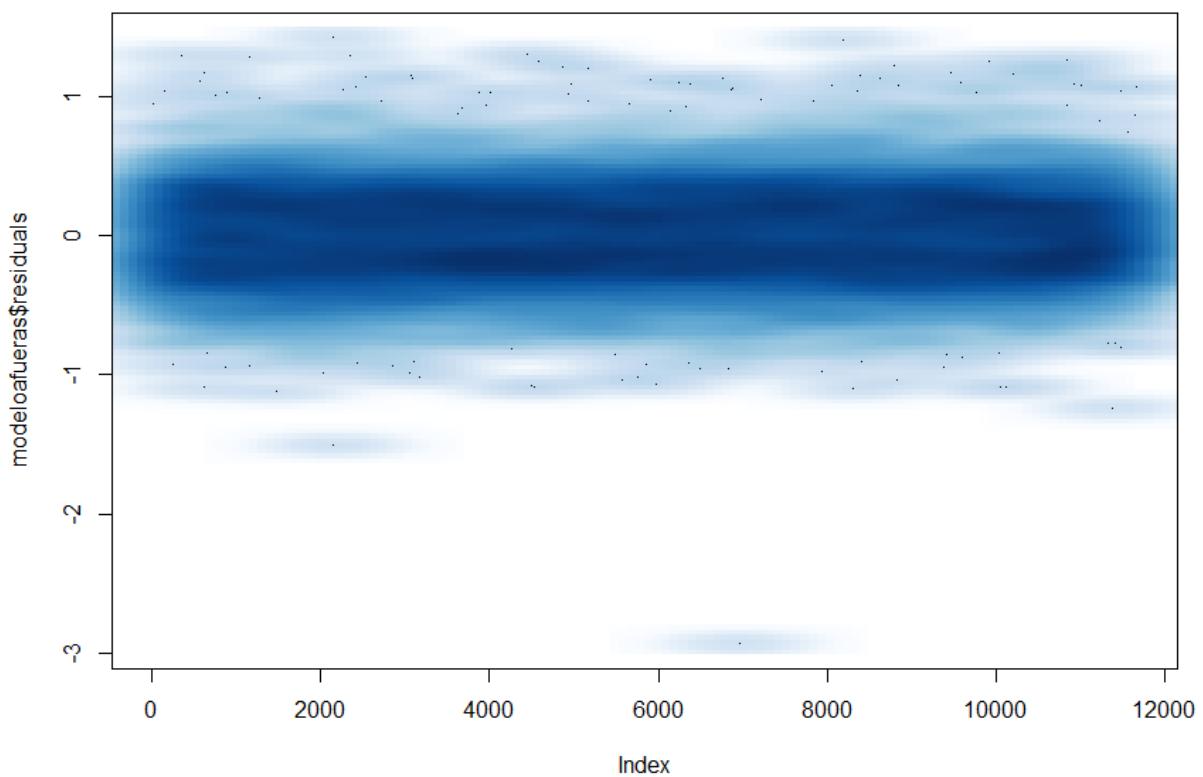
Los residuos tienen media cercana a cero y distribución parecida a normal:

Histogram of modeloafueras\$residuals



Normal Q-Q Plot





Voy a comprobar con modelo robusto si el modelo saldría mejor con un tratamiento de outliers.

```
Call: rlm(formula = log(price) ~ sqft_living, data = CASAS_AFUERAS)
Residuals:
    Min      1Q   Median      3Q     Max 
-2.972375 -0.228365 -0.003026  0.230003  1.416930 

Coefficients:
            value   Std. Error t value
(Intercept) 12.0393   0.0076 1582.1031
sqft_living  0.0004   0.0000 123.1880

Residual standard error: 0.3397 on 11680 degrees of freedom
```

Error en el modelo robusto aumenta ligeramente.

El intervalo de confianza es igual de estrecho como antes

	2.5 %	97.5 %
(Intercept)	1.202441e+01	1.205424e+01
sqft_living	4.050277e-04	4.181244e-04

Comparando la bondad de ambos modelos obtengo:

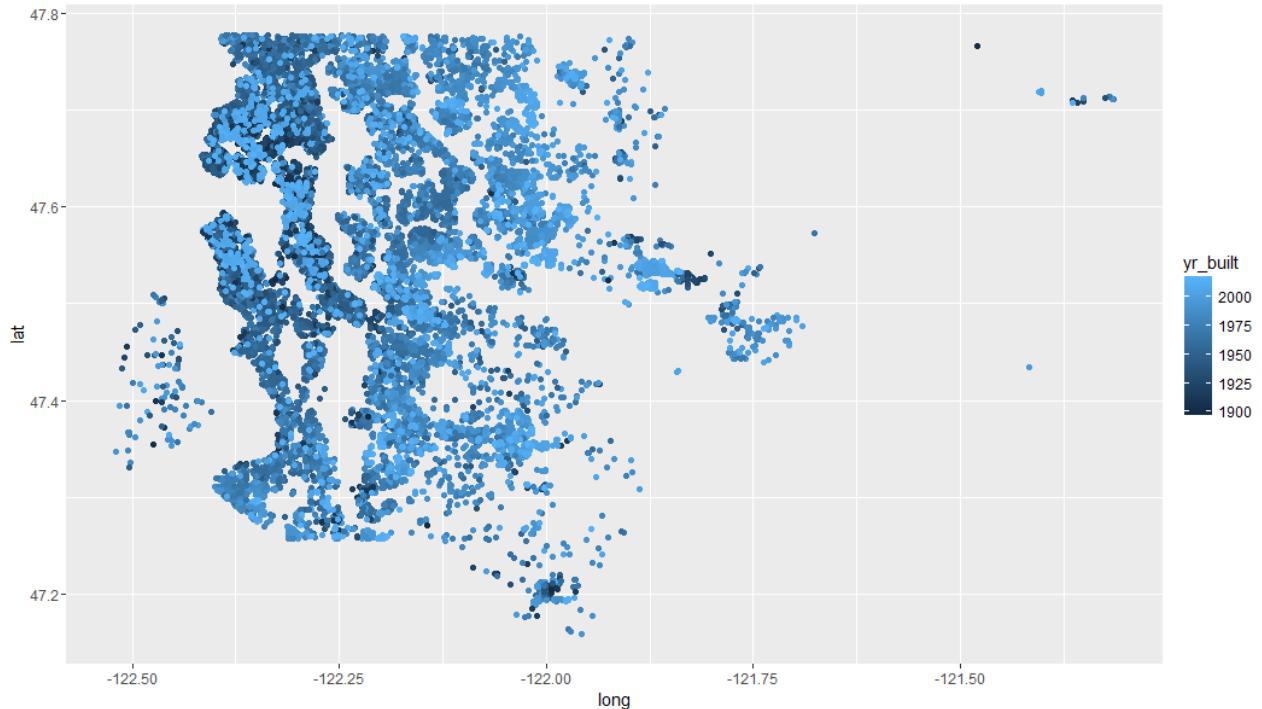
	modeloafueras	modeloafuerasrobusto
AIC	6691.435	6692.83
BIC	6713.532	6714.927

Concluyo que el modelo no mejora mediante tratamiento de outliers.

Análisis de las posibilidades de mejora de los modelos básicos mediante introducción de nuevas variables.

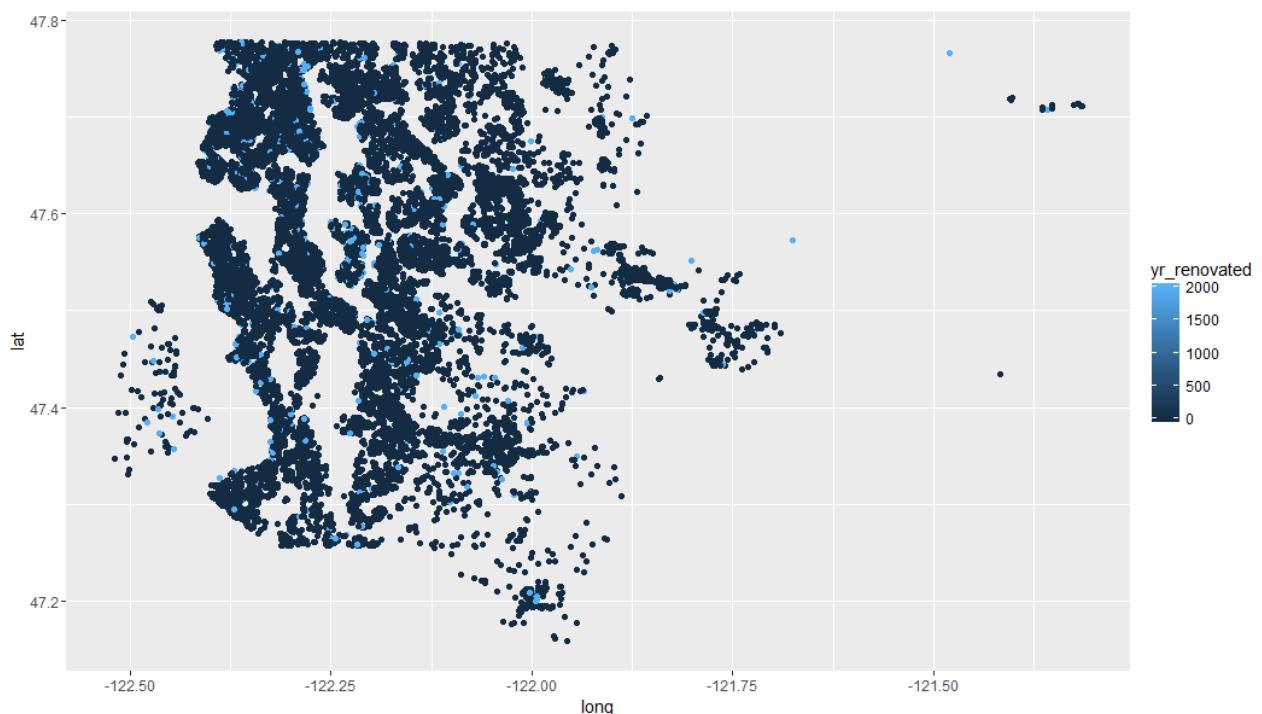
Tenemos disponibles distintas variables adicionales que podrían ayudar a mejorar el modelo básico. Voy a analizarlas:

La edad de las casas:



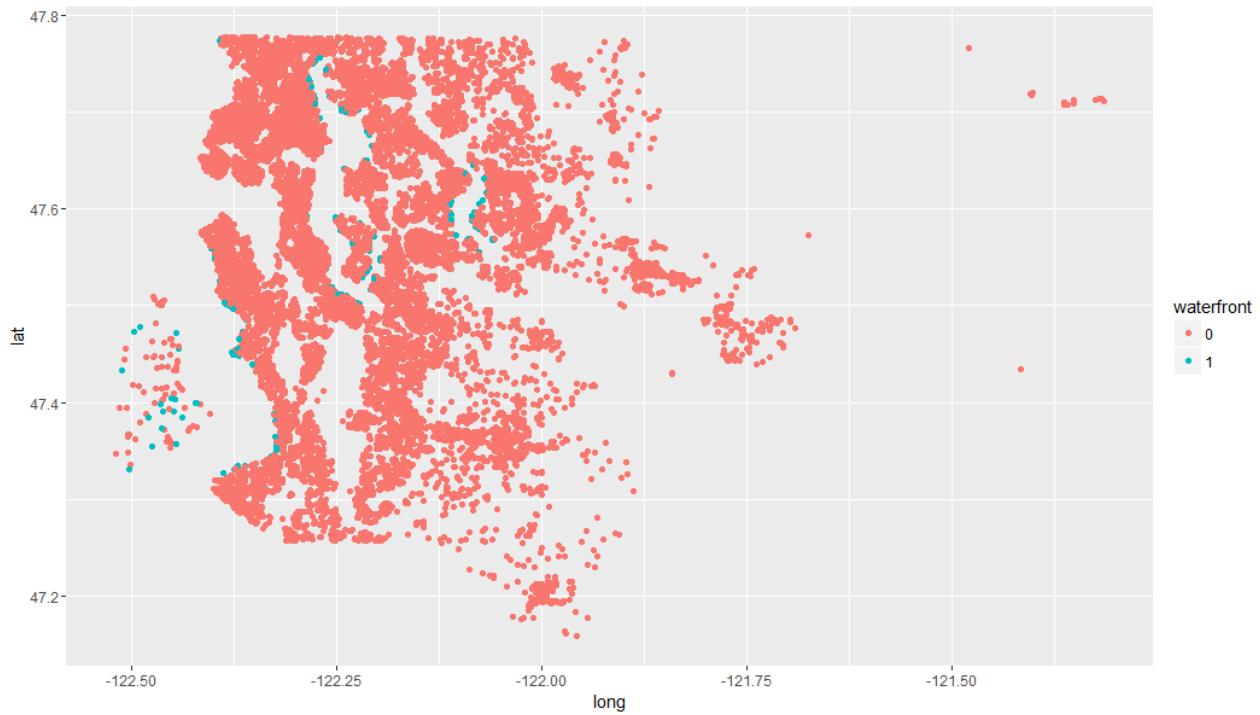
Las casas muy nuevas se encuentran en afuera, aunque también hay gran cantidad en el centro. En la cercanía a Capitol Hill se encuentran casas muy antiguas del principio del siglo XX.

Año de renovación:



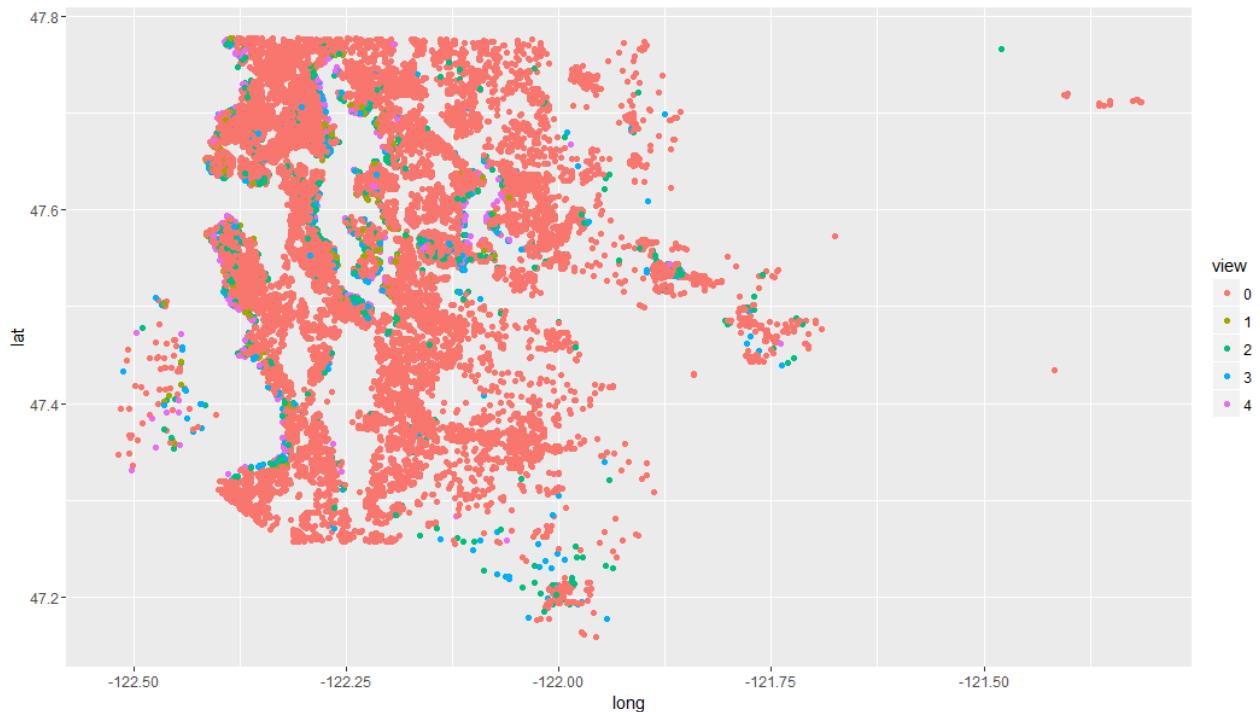
Bastantes casas en el centro y cerca del agua están renovadas. La mayoría de las casas de edad intermedia, construidas en los años 50-70 no parecen renovadas.

Vistas:



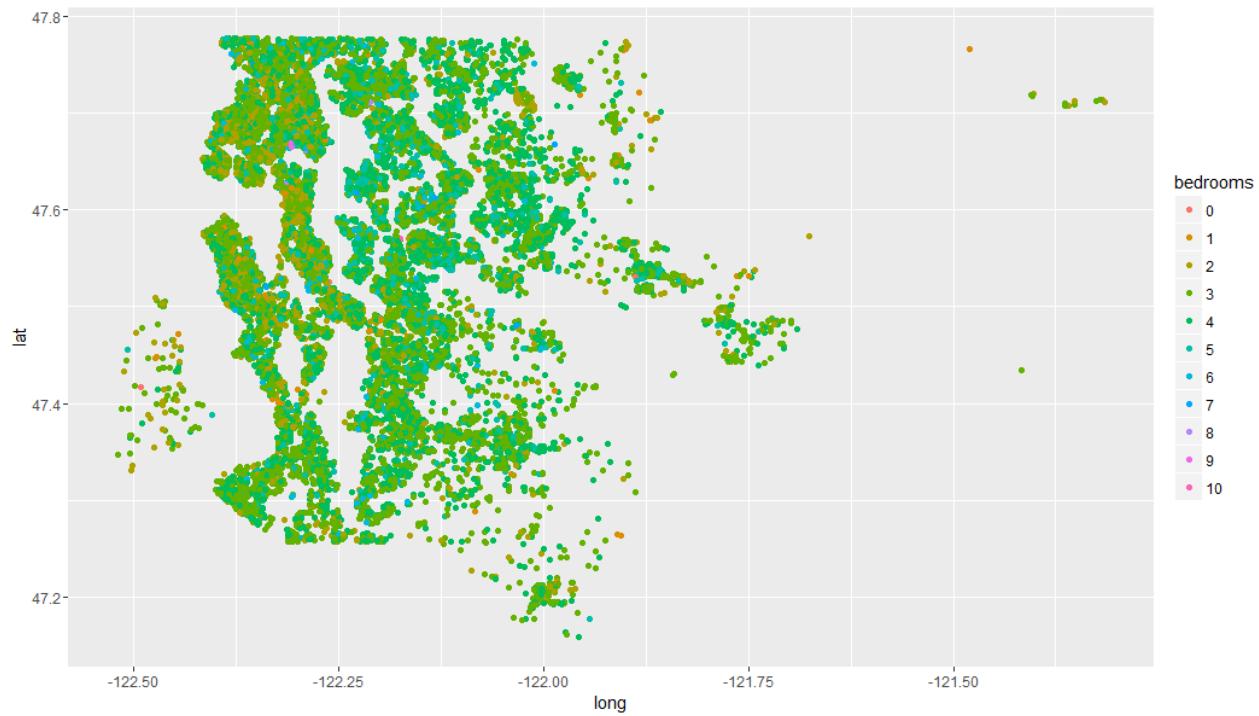
Cerca de la orilla hay casas en primera línea del mar con vistas en algunas zonas.

Orientación de la vivienda:



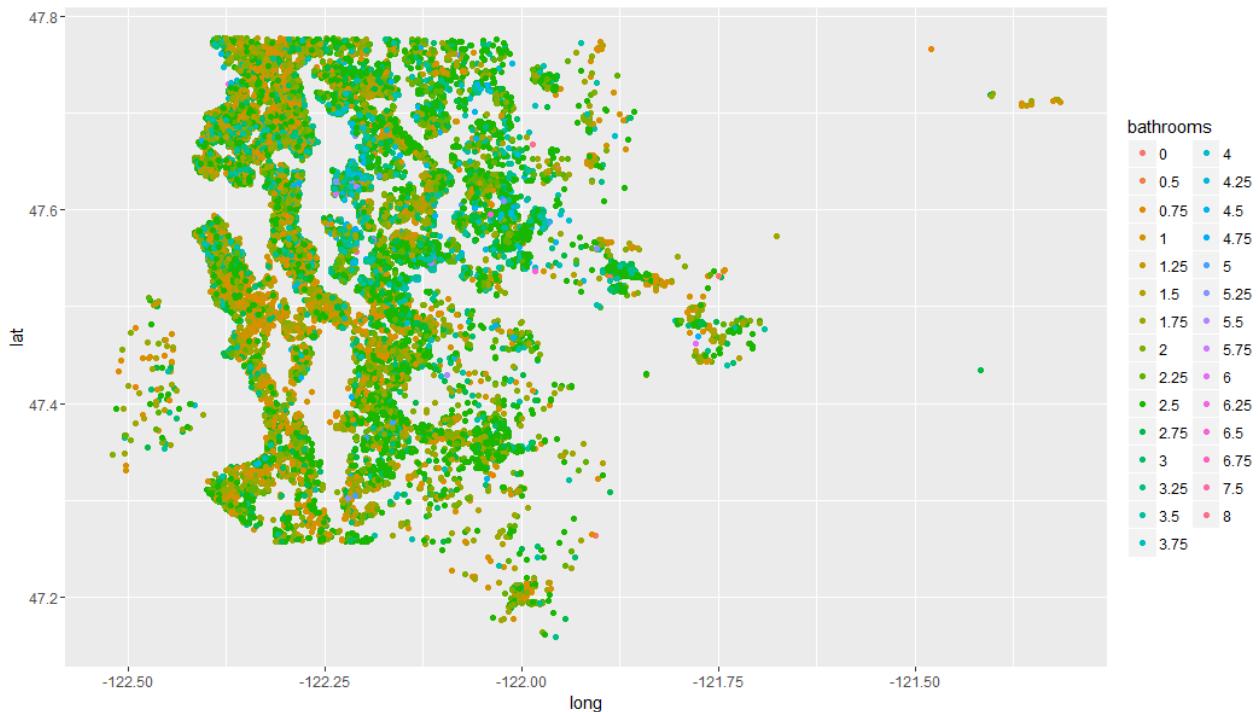
Cerca de la orilla las casas tienen también orientación mejor.

Número de habitaciones:



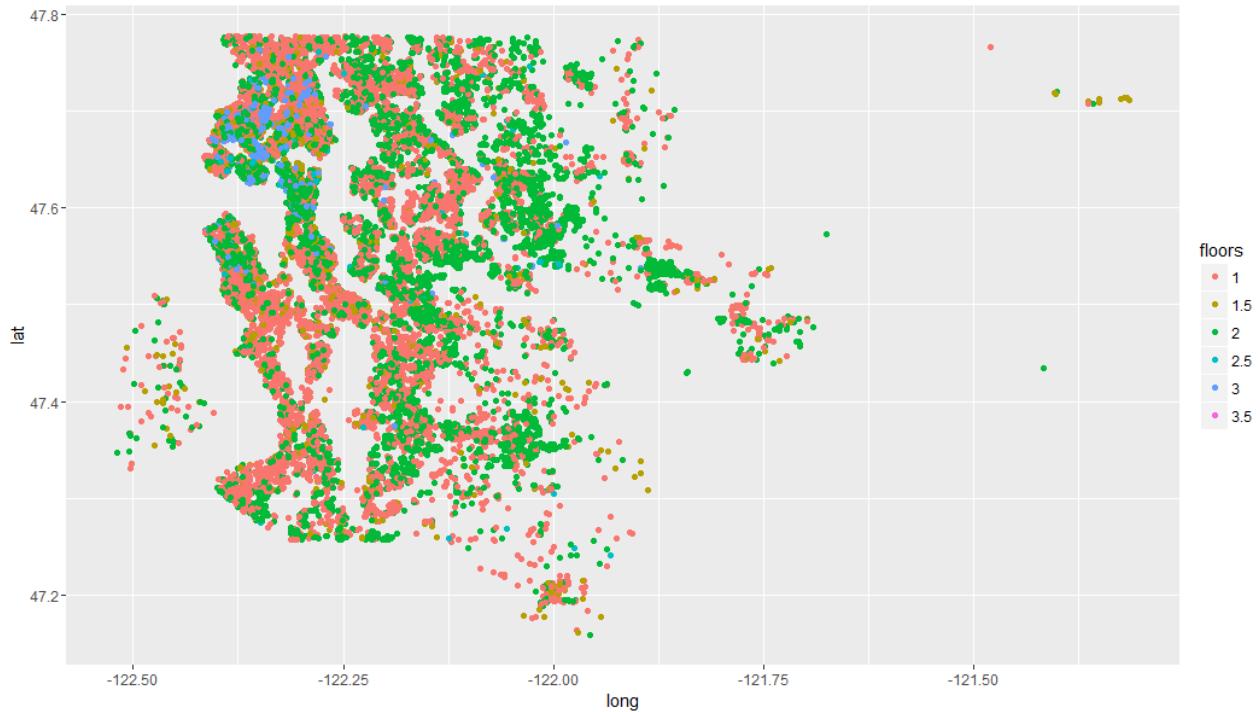
Las casas en el este parecen tener en media más habitaciones, lo cual coincide con el análisis del tamaño de las viviendas. Mayor número de 5-7 habitaciones observamos en la zona cara. También hay alguna casa de 10 habitaciones en el centro con vistas a Capitol Hill.

Número de baños:



Aunque no tengo claro que sería medio o cuarto de baño, puedo observar que las viviendas caras y grandes son donde hay más baños. Por lo contrario, las casas pequeñas tienen 1 baño o 2 como máximo por lo general.

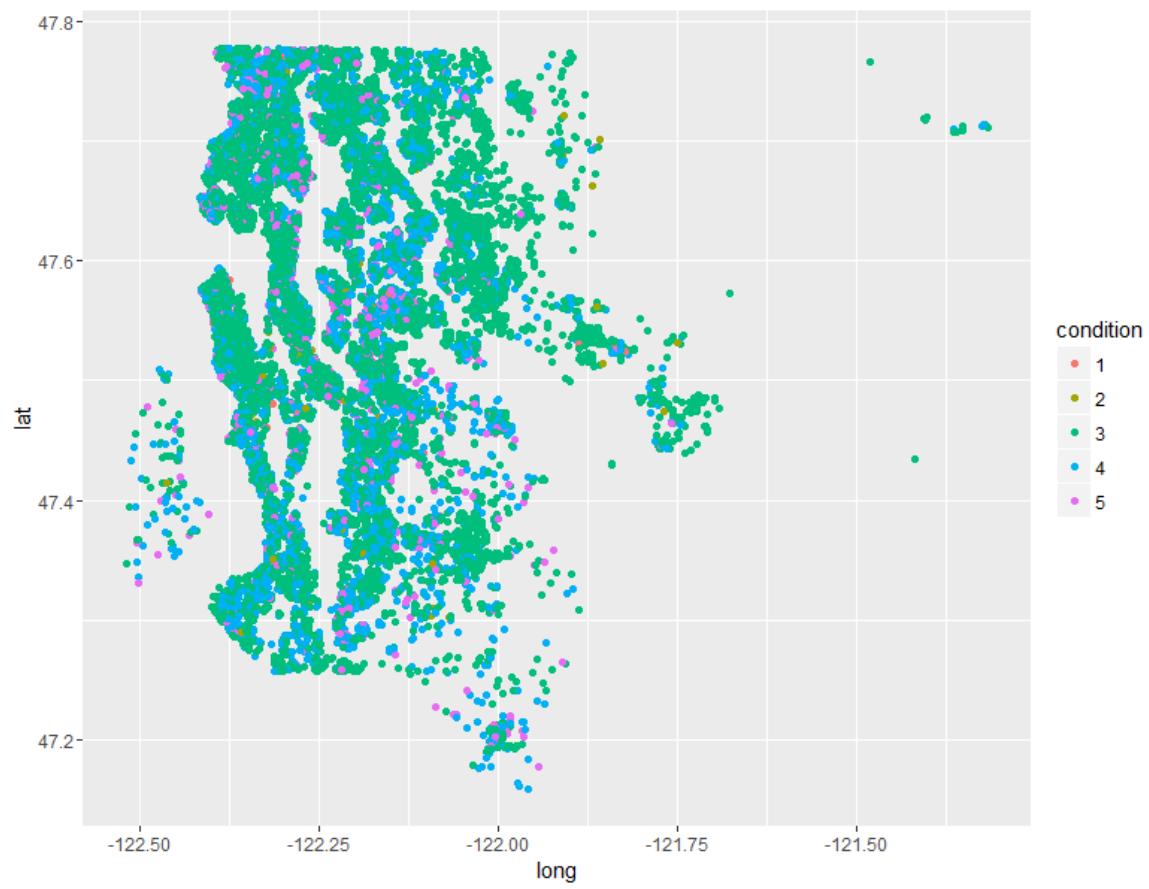
Número de plantas:



La mayoría de las viviendas tiene 1 o 2 plantas, excepto las viviendas en la zona norte de 3 plantas. Estas viviendas parecen ser las viviendas nuevas que se aprecian en el gráfico anterior.

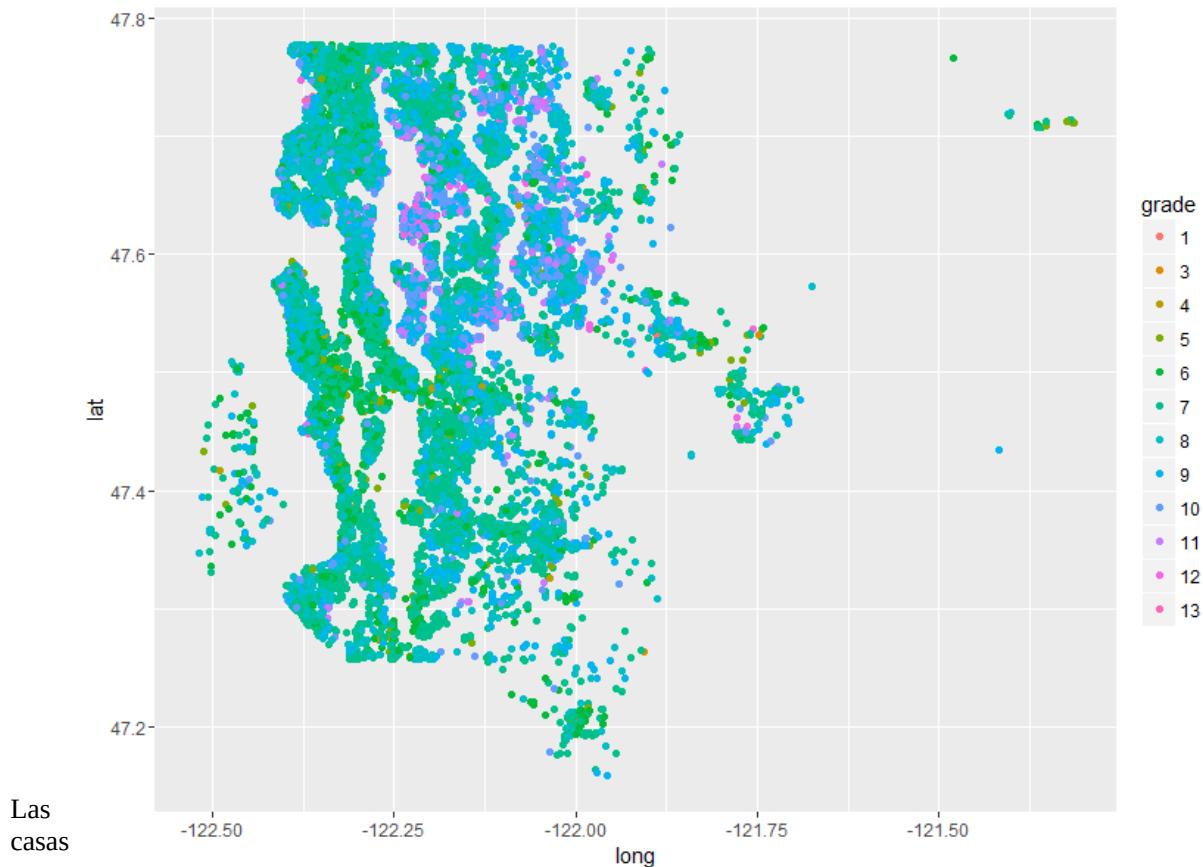
Las viviendas de 2 plantas en otras zonas también parecen ser más nuevas que las de 1 planta en la zona oeste.

Estado de la vivienda:



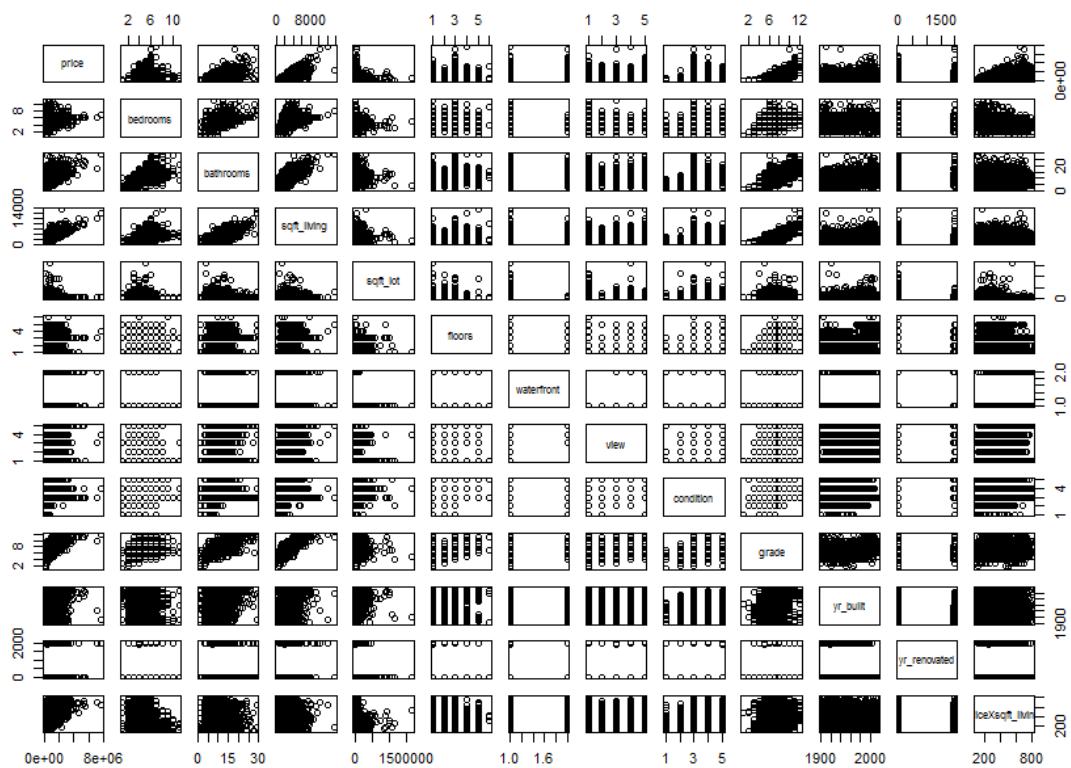
La mayoría de las viviendas está en medio o buen estado. Las casas peor mantenidas prevalecen en afueras.

Diseño:



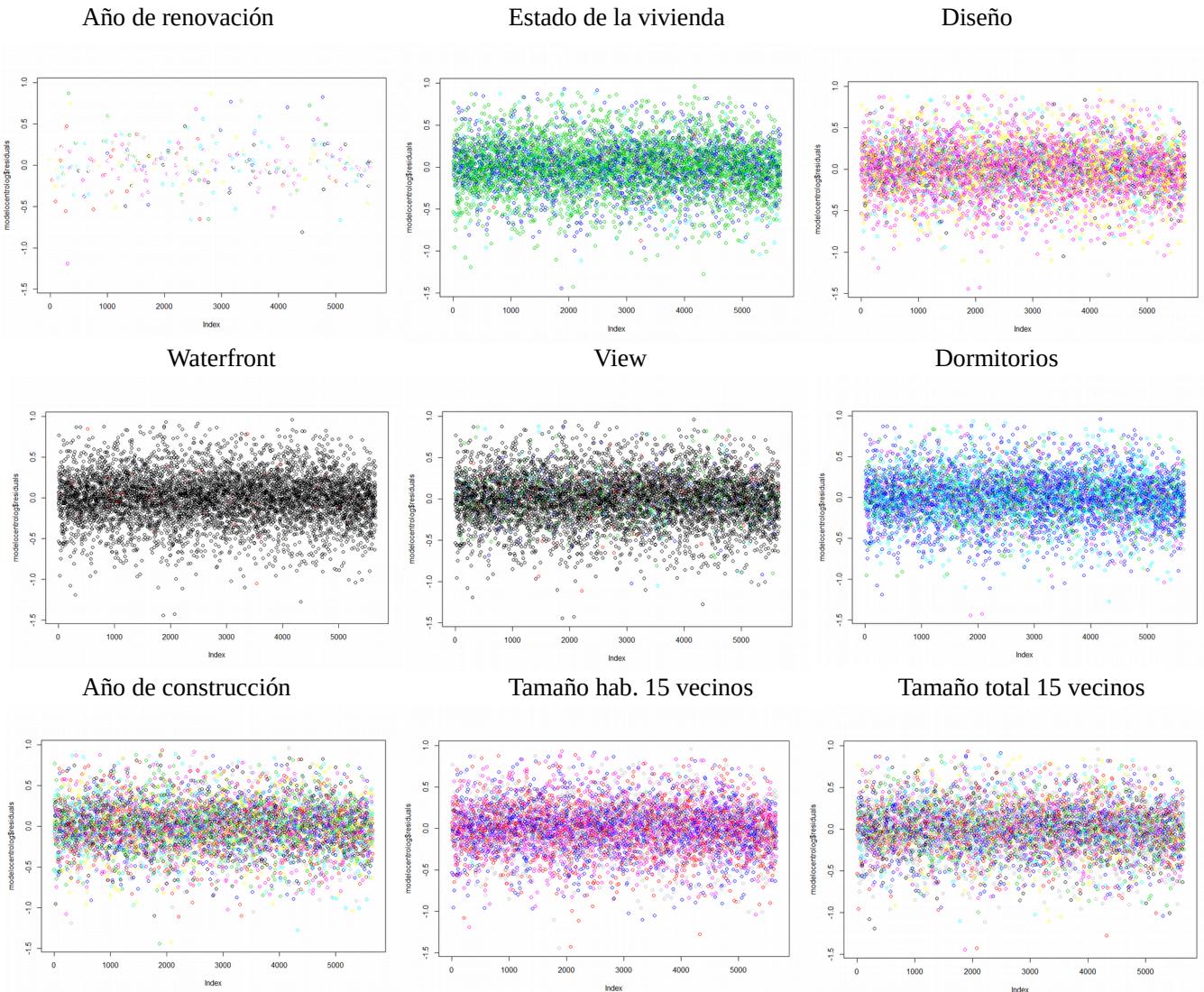
mejor diseñadas se encuentran definitivamente en la parte noreste, incluyendo los barrios caros y las zonas de viviendas nuevas. Las casas más sencillas están en la parte oeste de la ciudad.

Veamos la relación entre las variables de forma general:



Parece que cuanto más grande la vivienda (sqft_living) mas cara. La superficie parece tener relación también con el numero de dormitorios y de los baños. Cuanto mejor el diseño, mas baños y dormitorios por lo general. El precio por metro cuadrado es mayor cuando el diseño y estado de la casa mejora.

Voy a revisar si los residuos del modelo básico para las casas en el centro tienen algún patrón relacionado con estas variables.



No se ven patrones muy definidos aunque puede haber pequeñas asimetrías en el reparto de los residuos y la introducción de más variables puede ayudar a reducir el sesgo del modelo.

Veamos si con variables nuevas el modelo mejoraría. Construyo siguientes modelos añadiendo variables de firma incremental:

	AIC	BIC
Modelo básico	2182.275	2202.201
Añadiendo variable grade	1376.375	1456.08
Añadiendo variable condition	1086.48	1192.753
Añadiendo variable water front	1001.633	1114.548
Añadiendo variable view	857.4809	996.9642
Añadiendo var. año de constr.	318.1648	464.2902
Añadiendo var. sqft_living15	25.92166	178.6891

Me quedo con el modelo con AIC y BIC más bajo es el modelo que incluye las variables: sqft_living, grade, condition, waterfront, view, yr_built, sqft_living15.

El resultado parece bastante lógico: se paga más por cada unidad adicional de superficie en función del estado y diseño de la vivienda, dependiendo si tiene vistas y orientación mejor y también, dado que estamos hablando de la zona céntrica y cara que tiene edificios muy antiguos, en función de la edad de edificio y el vecindario que la rodea.

El mejor modelo para las casas en el centro que he conseguido es:

```
lm(formula = log(price) ~ sqft_living + grade + condition + waterfront +
view + yr_built + sqft_living15, data = CASAS_CENTRO)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.33243 -0.14826  0.00683  0.15254  1.00682 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.698e+01  2.766e-01 61.371 < 2e-16 ***
sqft_living  1.959e-04  5.719e-06 34.256 < 2e-16 ***
grade5       3.215e-01  1.472e-01  2.184  0.029023 *  
grade6       3.631e-01  1.405e-01  2.585  0.009763 ** 
grade7       5.755e-01  1.402e-01  4.107  4.07e-05 *** 
grade8       7.329e-01  1.403e-01  5.224  1.81e-07 *** 
grade9       9.414e-01  1.407e-01  6.689  2.46e-11 *** 
grade10      1.073e+00  1.416e-01  7.582  3.94e-14 *** 
grade11      1.123e+00  1.432e-01  7.837  5.49e-15 *** 
grade12      1.245e+00  1.486e-01  8.375 < 2e-16 *** 
grade13      1.098e+00  1.710e-01  6.424  1.44e-10 *** 
condition2   8.379e-02  7.959e-02  1.053  0.292536    
condition3   2.435e-01  7.132e-02  3.415  0.000643 *** 
condition4   2.775e-01  7.136e-02  3.889  0.000102 *** 
condition5   3.136e-01  7.177e-02  4.369  1.27e-05 *** 
waterfront1  2.740e-01  5.265e-02  5.205  2.01e-07 *** 
view1        7.222e-02  2.179e-02  3.315  0.000923 *** 
view2        4.105e-02  1.392e-02  2.950  0.003195 ** 
view3        6.944e-02  2.150e-02  3.230  0.001244 ** 
view4        1.943e-01  2.953e-02  6.580  5.11e-11 *** 
yr_builtin   -2.677e-03  1.166e-04 -22.959 < 2e-16 *** 
sqft_living15 1.271e-04  7.327e-06 17.345 < 2e-16 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.242 on 5643 degrees of freedom
```

Con el siguiente intervalo de confianza:

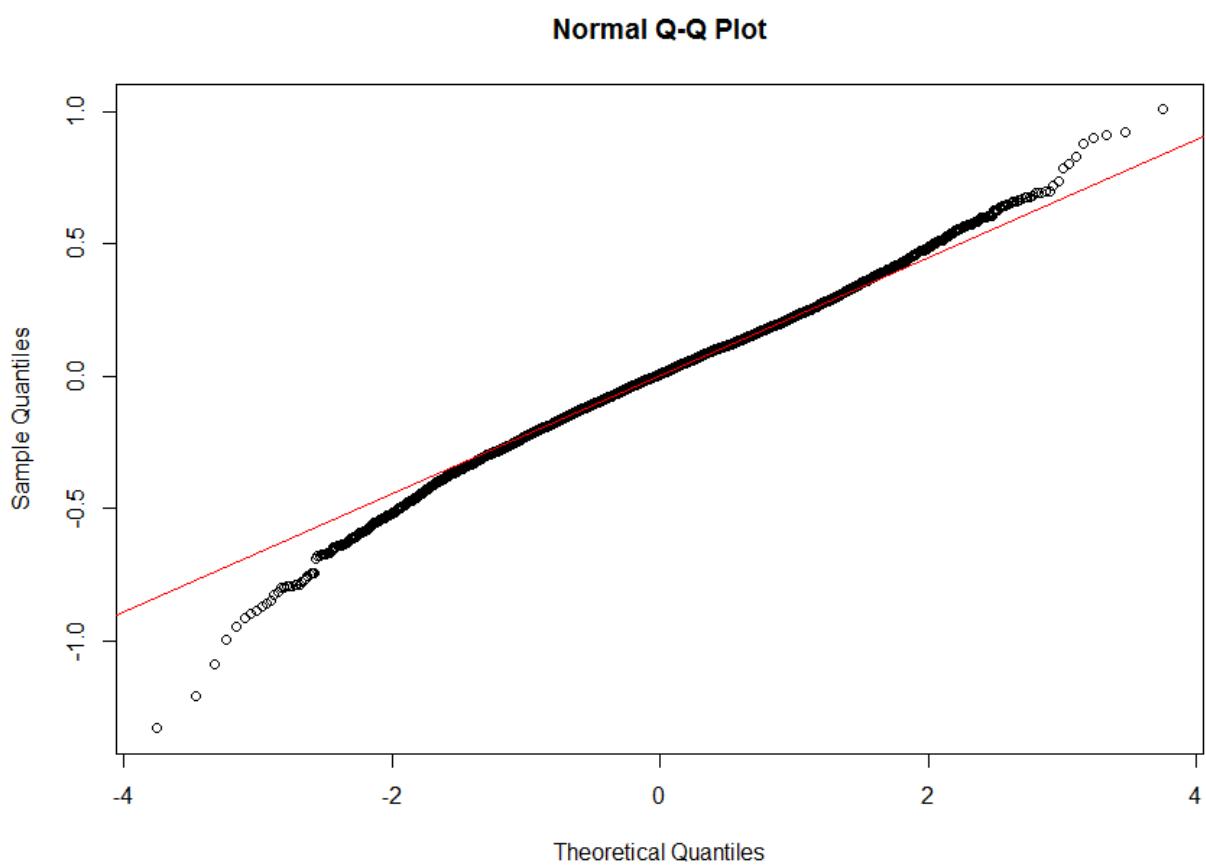
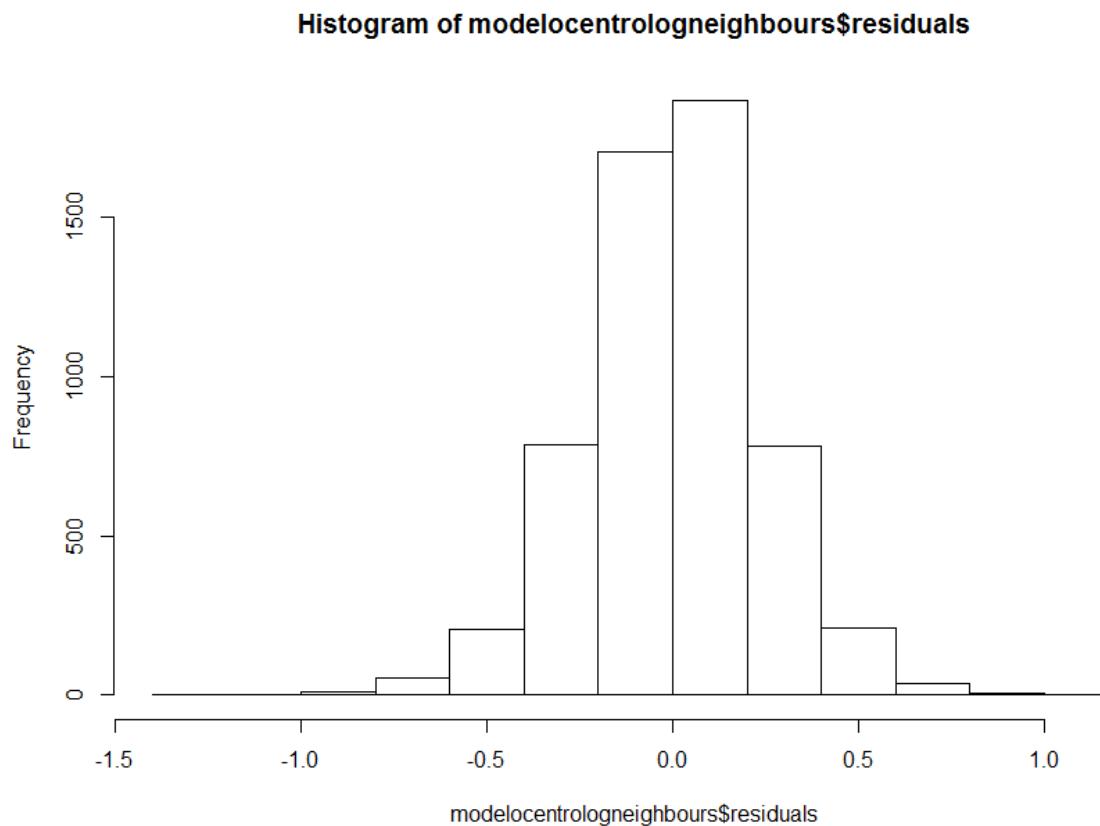
	2.5 %	97.5 %
(Intercept)	16.4335879268	17.518110479
sqft_living	0.0001846880	0.0002071095
grade5	0.0328812375	0.6100952464
grade6	0.0877329734	0.6384514031
grade7	0.3007875936	0.8502916036
grade8	0.4579005058	1.0079669452
grade9	0.6655338128	1.2173482038
grade10	0.7958754931	1.3509155356
grade11	0.8417191610	1.4033295296
grade12	0.9534345722	1.5361586251
grade13	0.7632224177	1.4336702745
condition2	-0.0722492597	0.2398238963
condition3	0.1037148719	0.3833565308
condition4	0.1376360546	0.4174134700
condition5	0.1728782714	0.4542642820
waterfront1	0.1708114925	0.3772340947
view1	0.0295073950	0.1149353998
view2	0.0137650389	0.0683257564
view3	0.0272949848	0.1115767329
view4	0.1364142421	0.2521800649
yr_builtin	-0.0029056617	-0.0024484902
sqft_living15	0.0001127211	0.0001414482

Es decir:

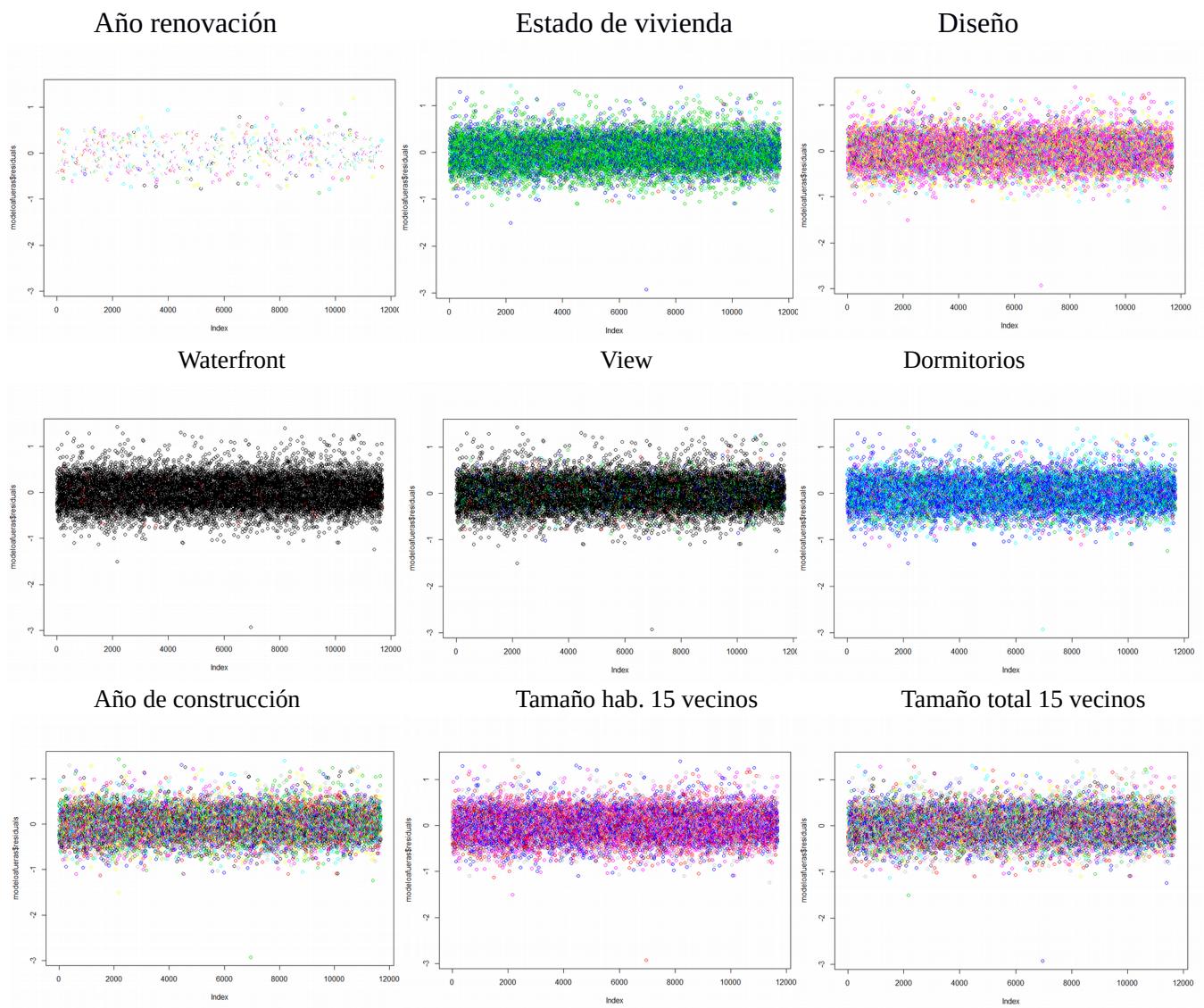
Con nivel de confianza de 95% cada pie cuadrado adicional de una vivienda de 2000 pies cuadrados de grade 13, condition 5, con vistas al mar y orientación (view) 4, construida en el año 2000 por ejemplo y con vecinos de casas en medio de 2000 pies cuadrados está asociado con un ascenso del precio entre:

(16,43+0,2+0,76+0,17+0,13-5,8+0,22) % = 12,28% y (17,52+0,4+1,4+0,45+0,38+0,25-4,8+0,28)% = 15,88%

Al añadir mas variables ha bajado el sesgo del modelo pero ha aumentado la dispersión.



Revisando los residuos del modelo básico de las casas en afueras:



No destacan patrones muy visibles. Aún así probaré mejorar el modelo con algunas variables adicionales y comprobando los resultados con AIC y BIC.

	AIC	BIC
Modelo básico	6691.435	6713.532
Añadiendo variable grade	4728.72	4831.841
Añadiendo variable condition	4564.302	4696.887
Añadiendo variable water front	3912.455	4052.405
Añadiendo variable view	3567.193	3736.606
Añadiendo variable yr_build	3500.594	3677.373
Añadiendo var. sqft_living15	3116.203	3300.348

El modelo con AIC y BIC más bajo incluye adicionalmente a la superficie las variables grade, condition, water front, view, yr_build y sqft_living15.

El mejor modelo para las casas de las afueras que he conseguido es:

```
lm(formula = log(price) ~ sqft_living + grade + condition + waterfront +
view + yr_built + sqft_living15, data = CASAS_AFUERAS)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.30240 -0.19647  0.00895  0.19840  1.26175 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.464e+01  3.970e-01 36.883 < 2e-16 ***
sqft_living 1.733e-04  5.281e-06 32.821 < 2e-16 ***
grade3      4.186e-03  3.299e-01  0.013  0.989879  
grade4      -9.394e-02  2.981e-01 -0.315  0.752633  
grade5      3.879e-02  2.888e-01  0.134  0.893159  
grade6      1.714e-01  2.887e-01  0.594  0.552721  
grade7      3.554e-01  2.888e-01  1.231  0.218372  
grade8      5.386e-01  2.888e-01  1.865  0.062260 .  
grade9      6.851e-01  2.890e-01  2.370  0.017789 *  
grade10     8.117e-01  2.893e-01  2.806  0.005029 ** 
grade11     8.821e-01  2.900e-01  3.041  0.002360 ** 
grade12     9.266e-01  2.926e-01  3.166  0.001547 ** 
grade13     1.188e+00  3.158e-01  3.762  0.000170 *** 
condition2  5.141e-02  8.865e-02  0.580  0.562006  
condition3  2.172e-01  8.409e-02  2.583  0.009797 ** 
condition4  2.227e-01  8.410e-02  2.648  0.008117 ** 
condition5  3.101e-01  8.456e-02  3.668  0.000246 *** 
waterfront1 4.834e-01  3.601e-02  13.423 < 2e-16 ***
view1       1.920e-01  2.457e-02  7.814  6.02e-15 *** 
view2       1.055e-01  1.348e-02  7.825  5.54e-15 *** 
view3       1.525e-01  1.802e-02  8.464 < 2e-16 *** 
view4       2.699e-01  2.895e-02  9.324 < 2e-16 *** 
yr_built    -1.550e-03  1.459e-04 -10.624 < 2e-16 *** 
sqft_living15 1.312e-04  6.628e-06 19.800 < 2e-16 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2762 on 11658 degrees of freedom
```

Con el siguiente intervalo de confianza:

	2.5 %	97.5 %
(Intercept)	13.8644725009	15.4208591015
sqft_living	0.0001629813	0.0001836853
grade3	-0.6425401547	0.6509111722
grade4	-0.6781940499	0.4903093378
grade5	-0.5273143004	0.6048936775
grade6	-0.3945411572	0.7373846690
grade7	-0.2105731534	0.9214640907
grade8	-0.0275998845	1.1047757292
grade9	0.1185431032	1.2516288575
grade10	0.2446080727	1.3787181997
grade11	0.3135838437	1.4505774201
grade12	0.3529985006	1.5001942400
grade13	0.5689625383	1.8070244083
condition2	-0.1223648907	0.2251798827
condition3	0.0524025494	0.3820597821
condition4	0.0578114035	0.3874972703
condition5	0.1443791991	0.4758818881
waterfront1	0.4127877733	0.5539627490
view1	0.1438250043	0.2401475609
view2	0.0790639646	0.1319184373
view3	0.1172169597	0.1878704047
view4	0.2131836635	0.3266753942
yr_built	-0.0018355517	-0.0012637241
sqft_living15	0.0001182366	0.0001442191

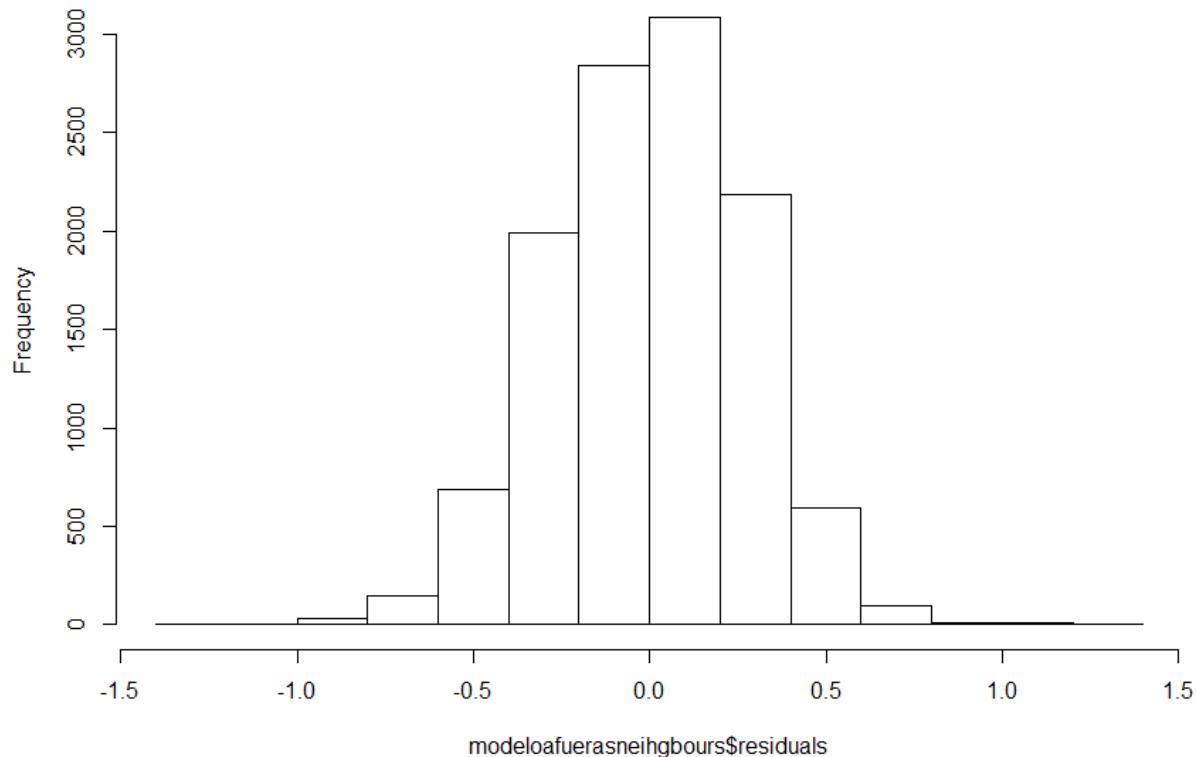
Es decir:

Con nivel de confianza de 95% cada pie cuadrado adicional de una vivienda de 2000 pies cuadrados de grade 13, condition 5, con vistas al mar y orientación (view) 4 y con vecinos de casas en medio de 2000 pies cuadrados constuido en el año 2000 está asociado con un ascenso del precio entre:

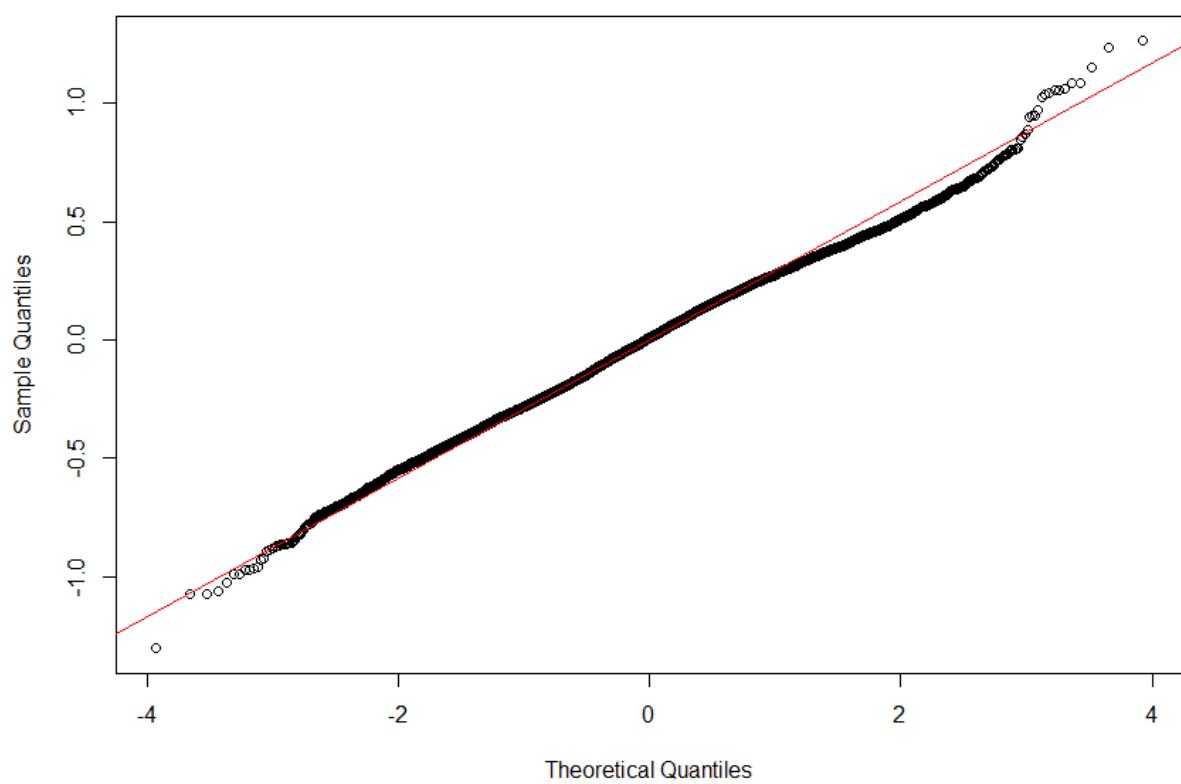
$(13.86+0.32+0.57+0.144+0.4+0.21-3.6+0.22)\% = 12.12\%$ y $(15.42+0.39+1.8+0.47+0.55+0.33-2.5+0.28)\% = 16.83\%$

Al añadir mas variables ha bajado el sesgo del modelo pero ha aumentado la dispersión.

Histogram of modeloafuerasneihgbours\$residuals



Normal Q-Q Plot



Estimación el precio de venta de los inmuebles de la cartera de la empresa.