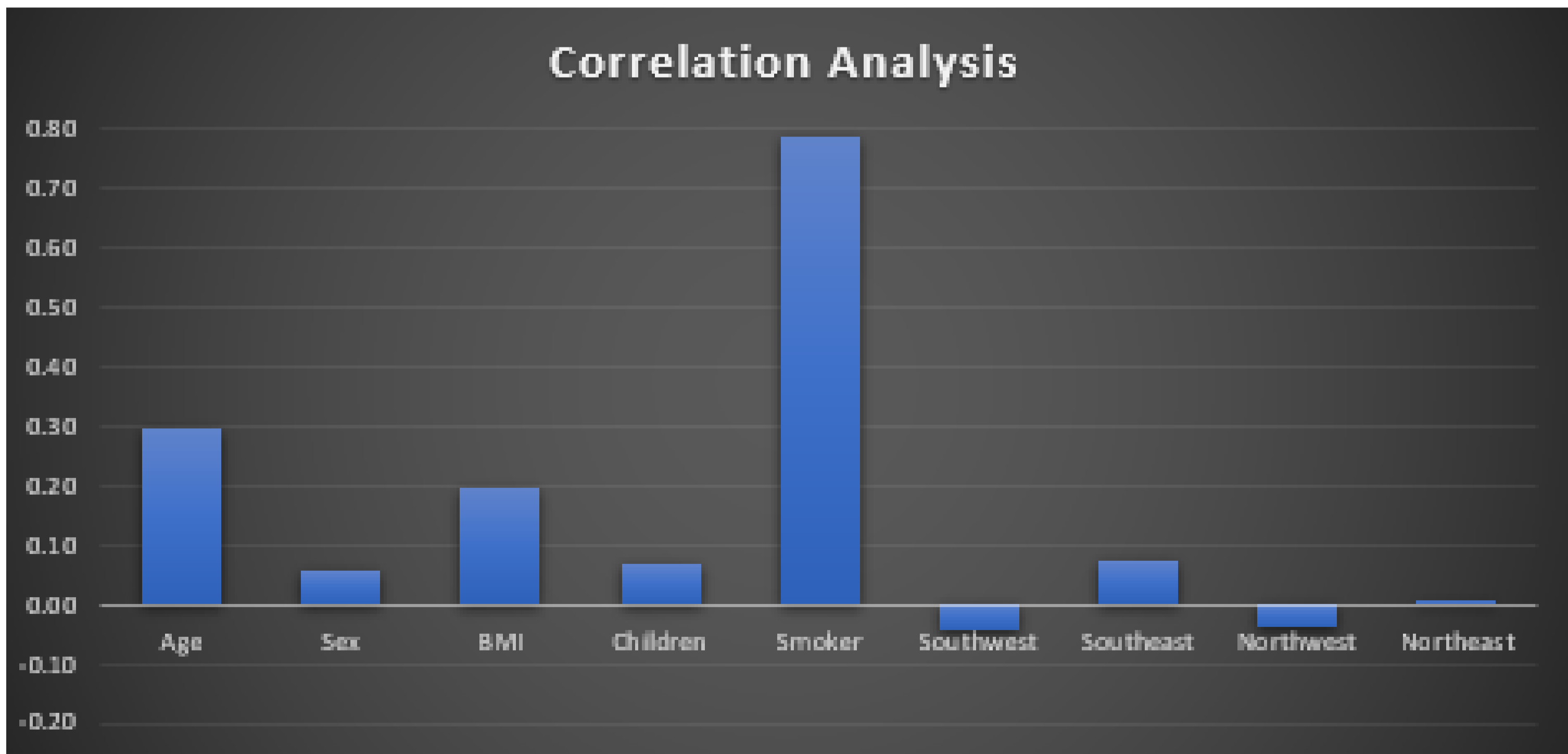


- **Age:** The sample set includes individuals who are 18 years old. All the records in the sample set have an age value of 18.
- **Sex:** The sex of the individuals is indicated by the variable "M1 F0." The value "1" represents male, and "0" represents female. In the sample set, there are both males and females.
- **BMI:** The sample set includes different values for the body mass index (BMI) of the individuals. The BMI values range from 17.29 to 42.24, indicating variations in weight status.
- **Children:** The number of children is represented by the variable "Children." The sample set contains individuals with different numbers of children, ranging from 0 to 4.
- **Smoker:** The variable "Smoker" indicates whether an individual is a smoker or not. The value "1" represents smokers, and "0" represents non-smokers. In the sample set, there are both smokers and non-smokers.
- **Region:** The variables "Southwest," "Southeast," "Northwest," and "Northeast" represent the different regions of the individuals. Each region is indicated by a binary value of "1" or "0." The sample set includes individuals from multiple regions, with varying combinations of region values.
- **Charges:** The charges variable represents the medical insurance charges for each individual. The charges vary widely across the sample set, ranging from \$2,801.26 to \$38,792.69.

	Age	Sex	BMI	Children	Smoker	Southwest	Southeast	Northwest	Northeast	Charges
Age	1.00	-0.02	0.11	0.04	-0.03	0.01	-0.01	0.00	0.00	0.30
Sex	-0.02	1.00	0.05	0.02	0.08	0.00	0.02	-0.01	0.00	0.06
BMI	0.11	-0.02	1.00	0.01	0.00	-0.01	0.27	-0.14	-0.14	0.20
Children	0.04	0.02	0.01	1.00	0.01	0.02	-0.02	0.02	-0.02	0.07
Smoker	-0.03	0.08	0.00	0.01	1.00	-0.04	0.07	-0.04	0.00	0.79
Southwest	0.01	0.00	-0.01	0.02	-0.04	1.00	-0.35	-0.32	-0.32	-0.04
Southeast	-0.01	0.02	0.27	-0.02	0.07	-0.35	1.00	-0.35	-0.35	0.07
Northwest	0.00	-0.01	-0.14	0.02	-0.04	-0.32	-0.35	1.00	-0.32	-0.04
Northeast	0.00	0.00	-0.14	-0.02	0.00	-0.32	-0.35	-0.32	1.00	0.01
Charges	0.30	0.06	0.20	0.07	0.79	-0.04	0.07	-0.04	0.01	1.00



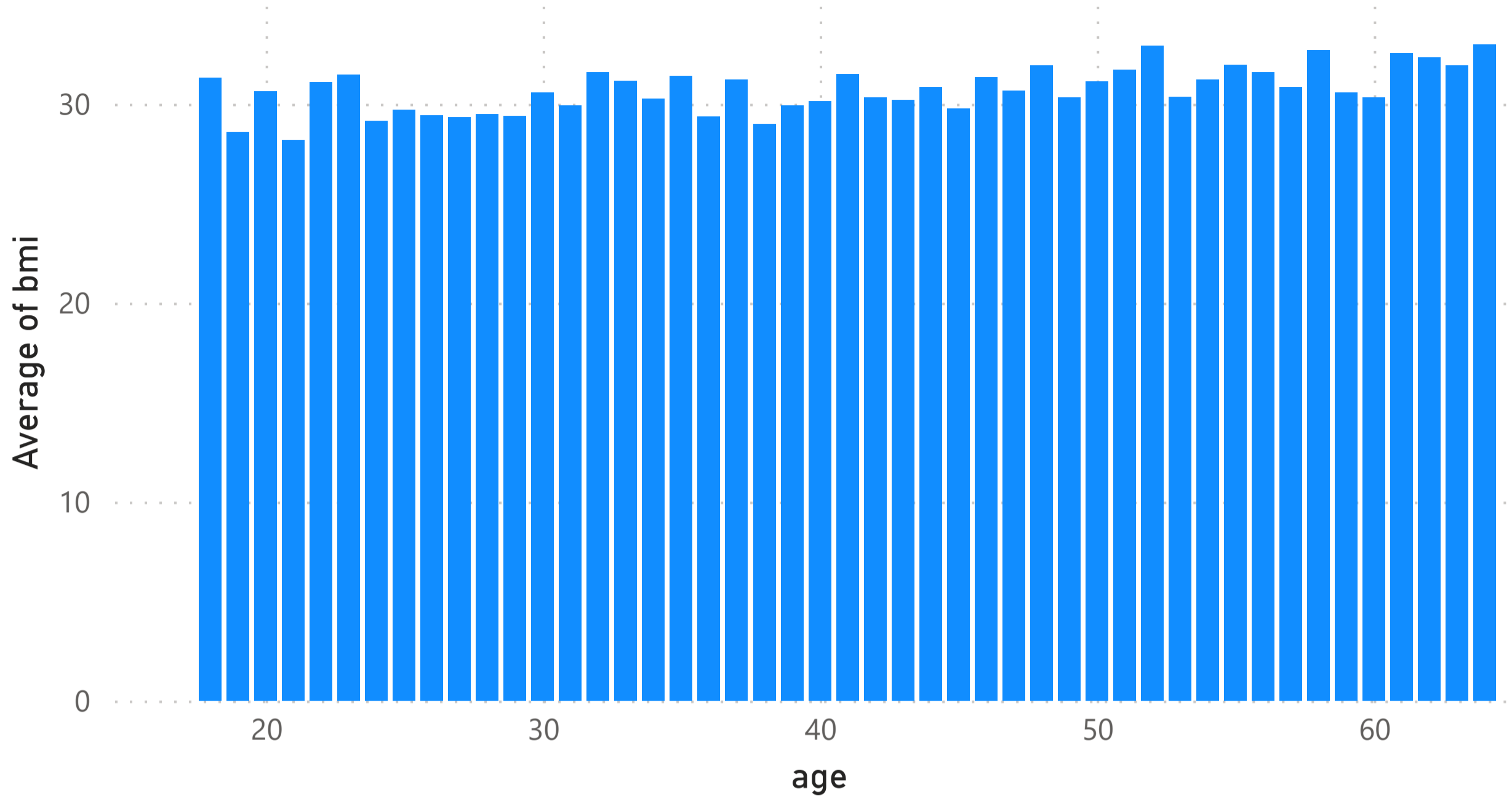
Age: There is a positive correlation of 0.30 between age and charges. This suggests that as age increases, medical insurance charges tend to increase. It indicates that older individuals may have higher medical expenses, leading to higher insurance charges.

Smoker: There is a strong positive correlation of 0.79 between being a smoker and charges. This indicates that smokers tend to have significantly higher medical insurance charges compared to non-smokers. It suggests that smoking is a significant factor contributing to increased healthcare costs.

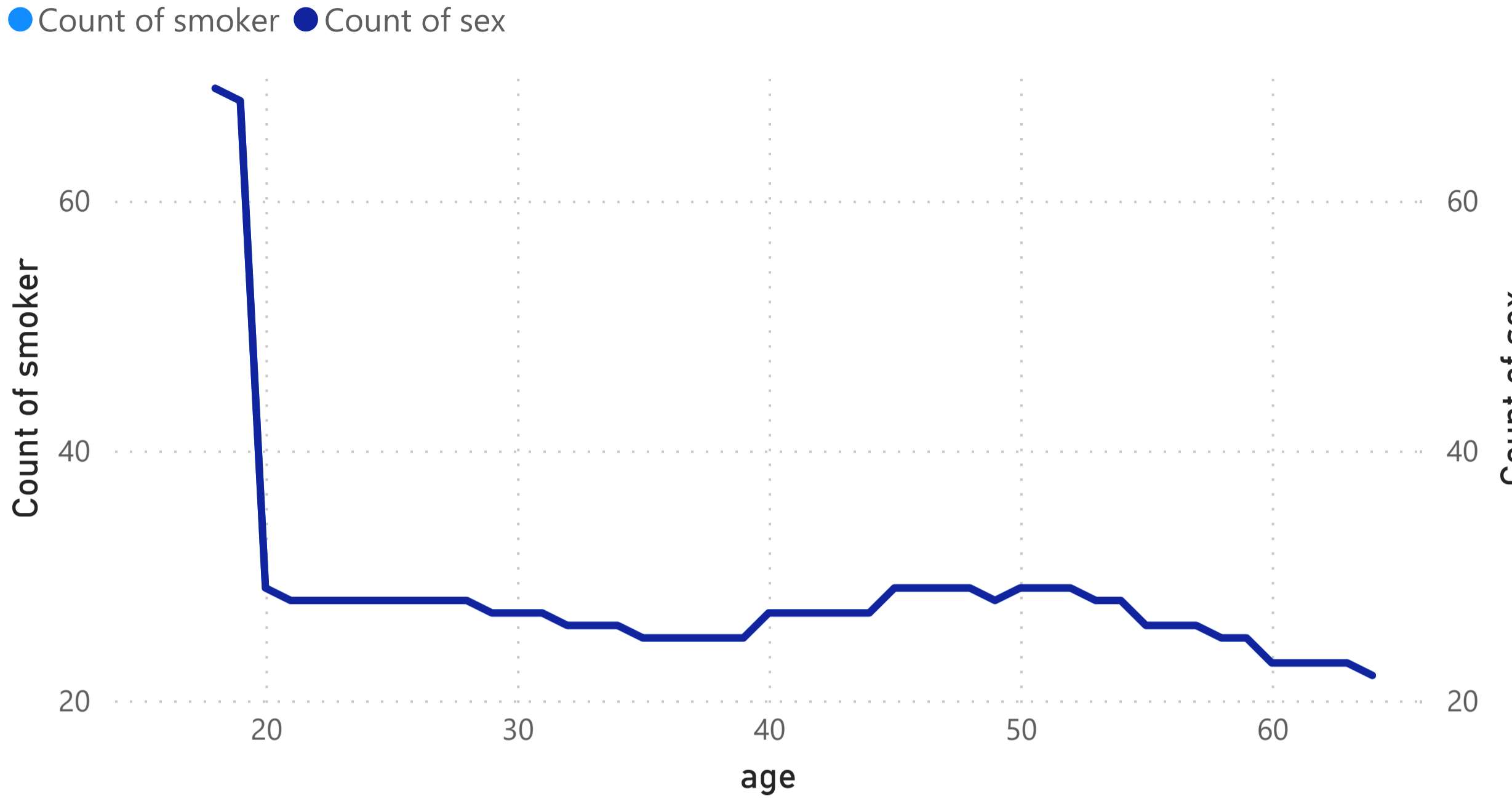
BMI: There is a moderate positive correlation of 0.20 between BMI and charges. This suggests that individuals with higher body mass index (BMI) values tend to have slightly higher medical insurance charges. It implies that higher BMI may be associated with increased health risks and medical expenses.

Region: The variables Southwest, Southeast, Northwest, and Northeast represent different regions. However, their correlations with charges are relatively weak and range between -0.04 and 0.01. This implies that the region of residence has limited impact on medical insurance charges, at least based on the linear relationship captured by the correlation coefficients.

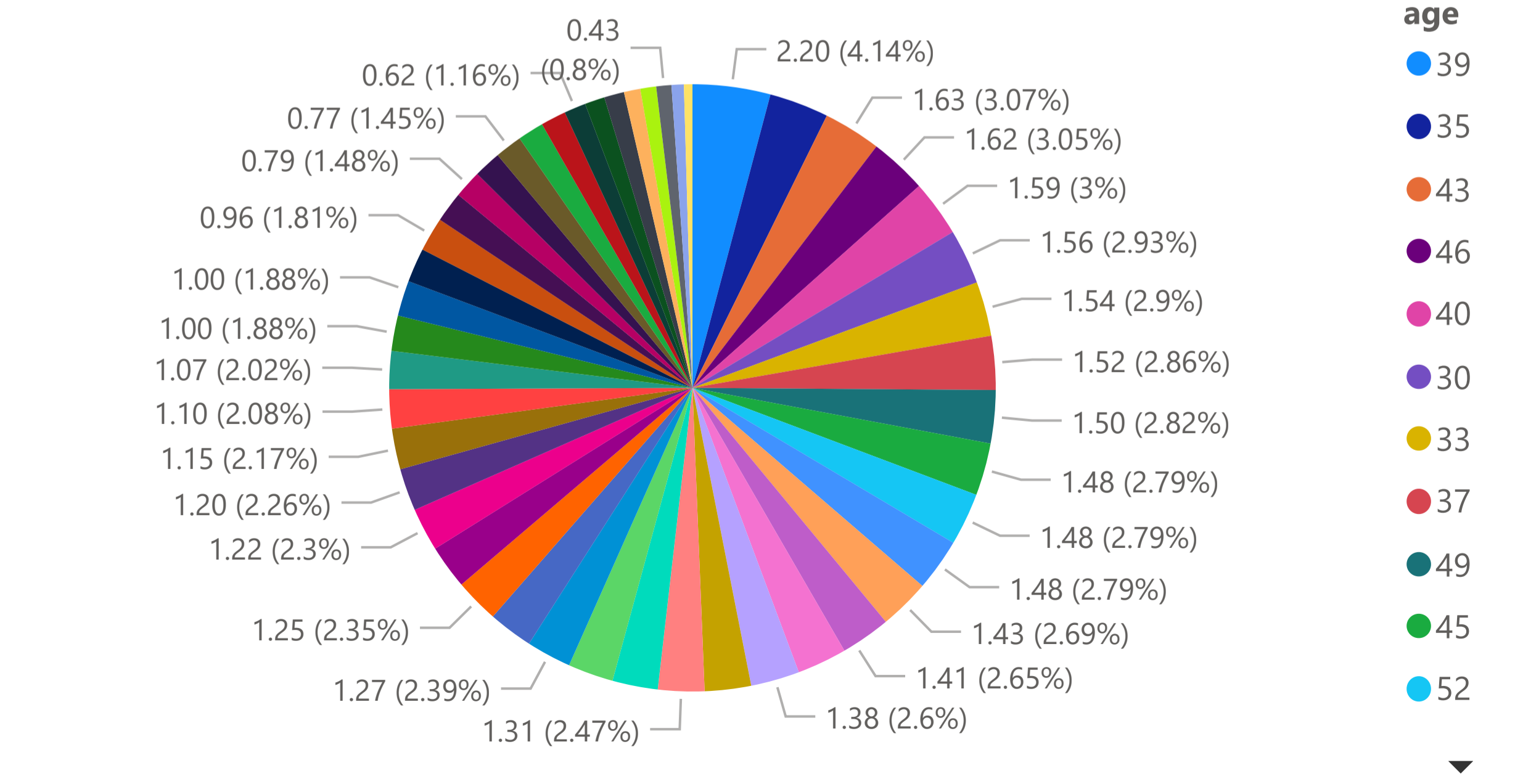
Average of bmi by age



Count of smoker and Count of sex by age



Average of children by age



AGE vs BMI: At 32.98, 64 had the highest Average of bmi and was 17.00% higher than 21, which had the lowest Average of bmi at 28.19. Across all 47 age, Average of bmi ranged from 28.19 to 32.98.

Age vs Children: 19 had the highest Average of children at 2.20, followed by 35 and 43. 58 had the lowest Average of children at .24.

Age Vs Smoker & Sex: At 69, 18 had the highest Count of smoker and was 213.64% higher than 64, which had the lowest Count of smoker at 22.

Count of smoker and total Count of sex are positively correlated with each other.

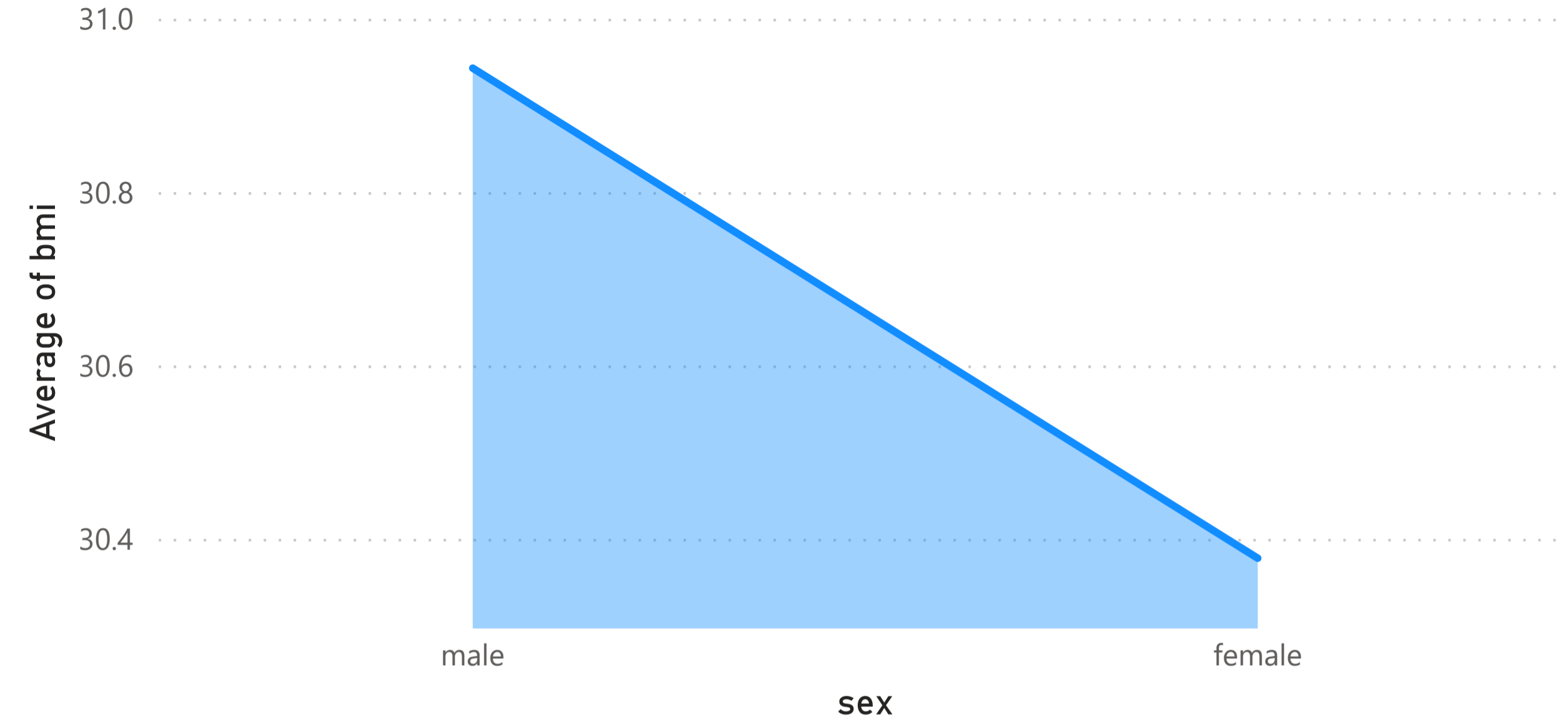
18 accounted for 5.16% of Count of smoker.

Age Vs BMI: 52 had the highest Average of bmi (32.94). 51 and 52 tied for the highest Count of bmi (29).

BMI vs Sex: Average of bmi for male (30.94) was higher than female (30.38). female had 30.38 Average of bmi and male had 30.94.

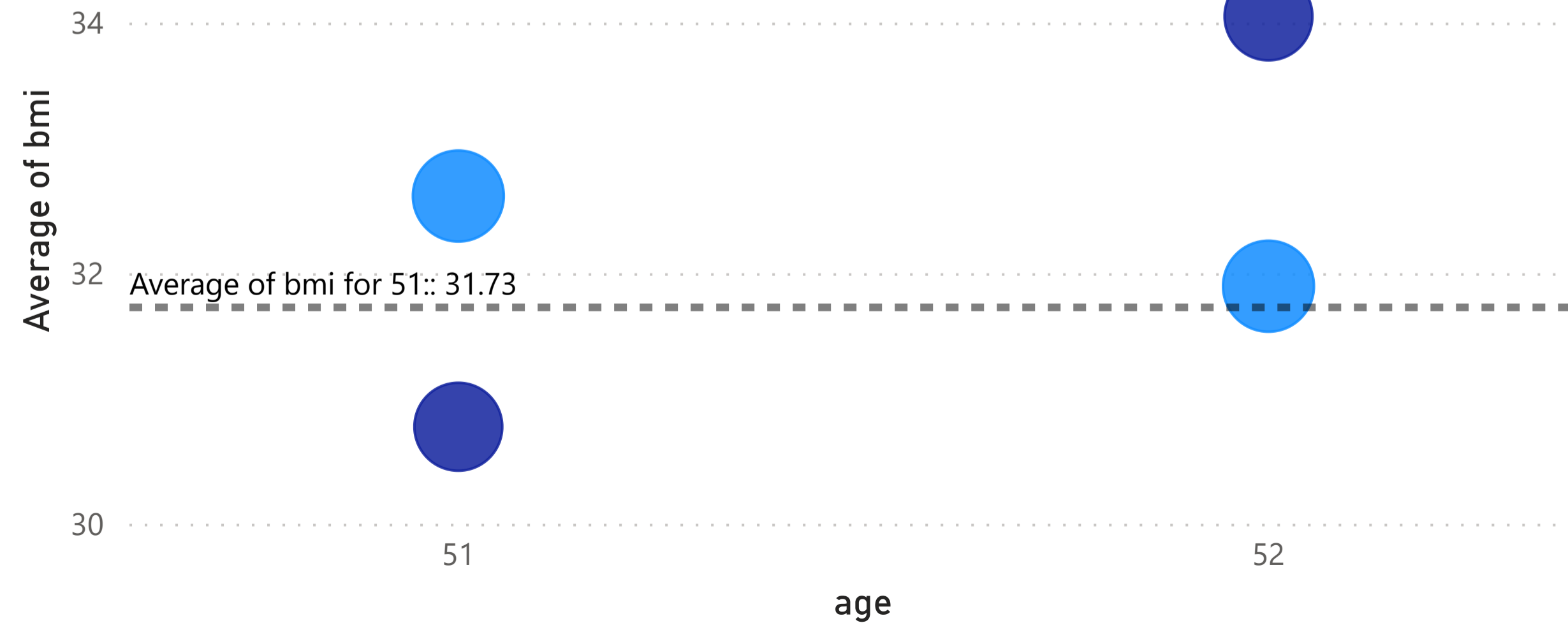
BMI vs Smoker: Count of smoker was highest for 32.30 at 13, followed by 28.31 and 30.80. 32.30 accounted for 0.97% of Count of smoker. Across all 548 bmi, Count of smoker ranged from 1 to 13.

Average of bmi by sex

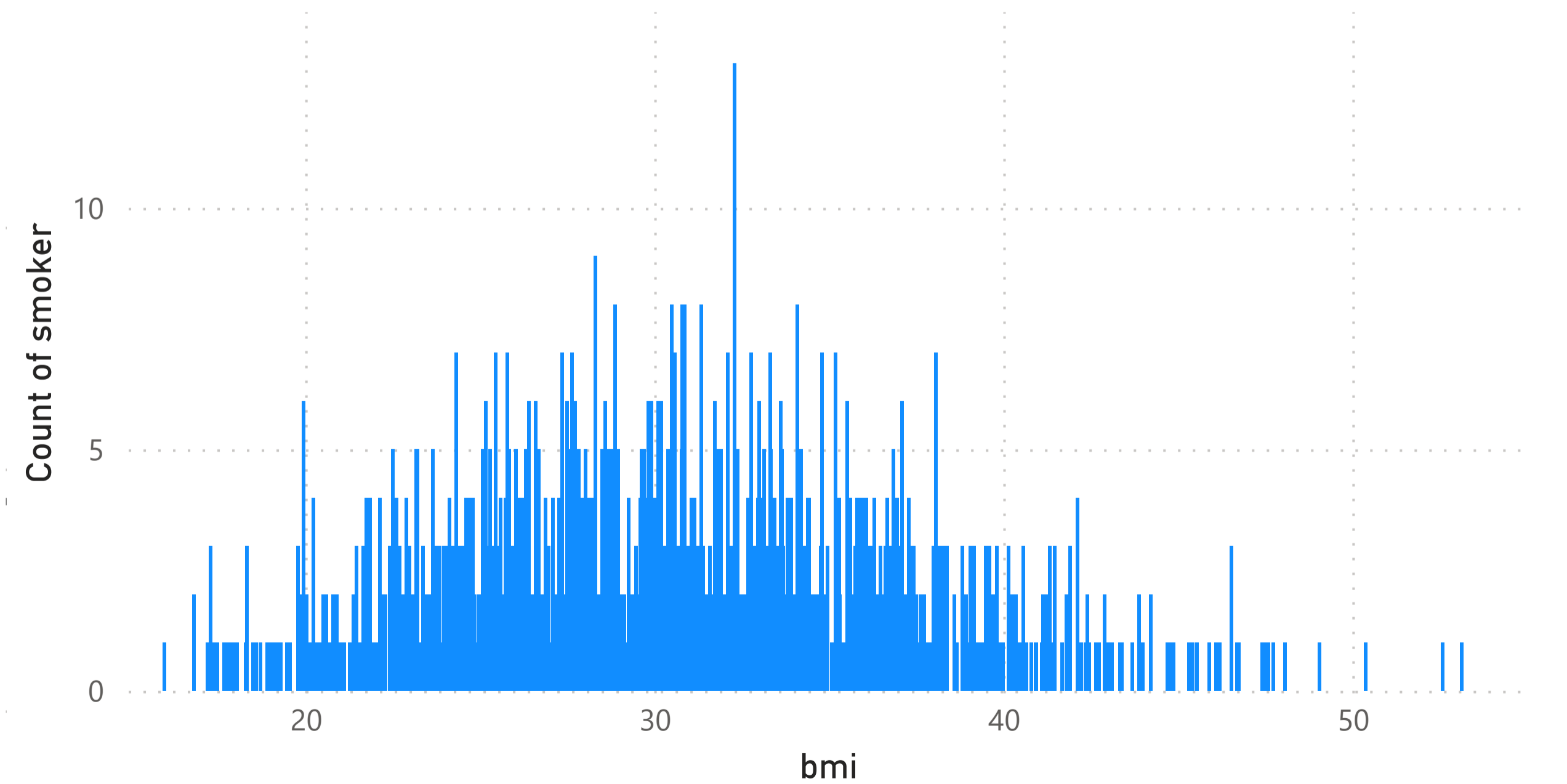


Average of bmi and Count of bmi by sex and age

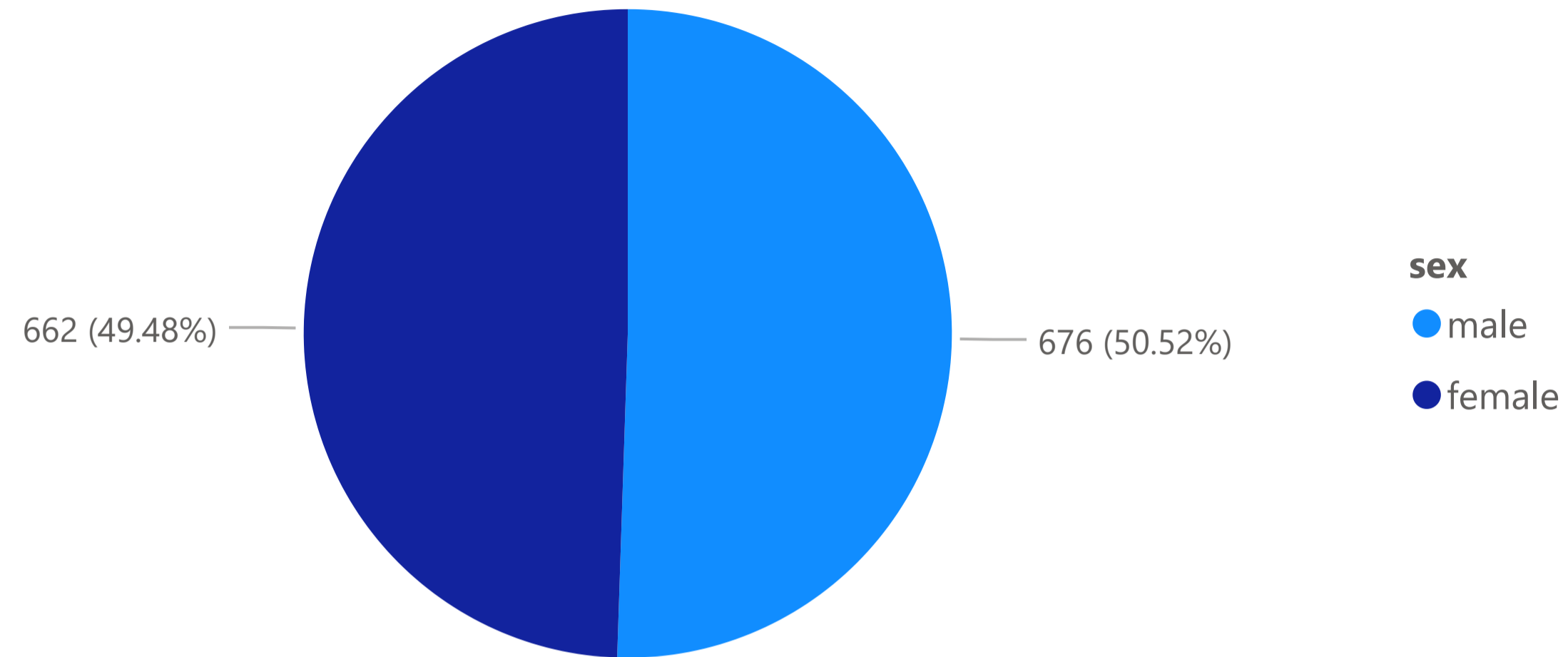
sex ● female ● male



Count of smoker by bmi



Count of smoker by sex

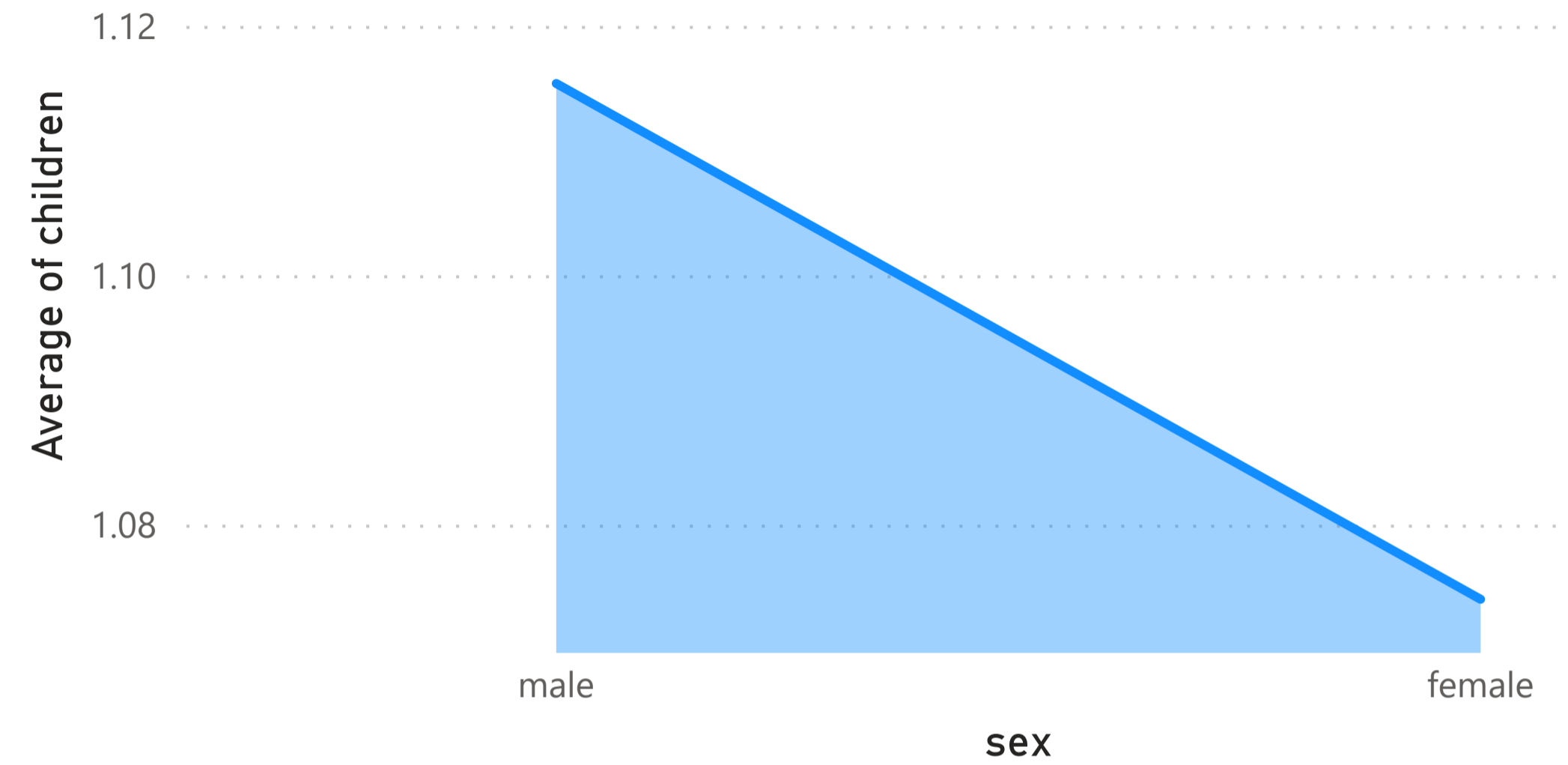


Sex Vs Smoker: Count of smoker for male (676) was higher than female (662). male accounted for 50.52% of Count of smoker.

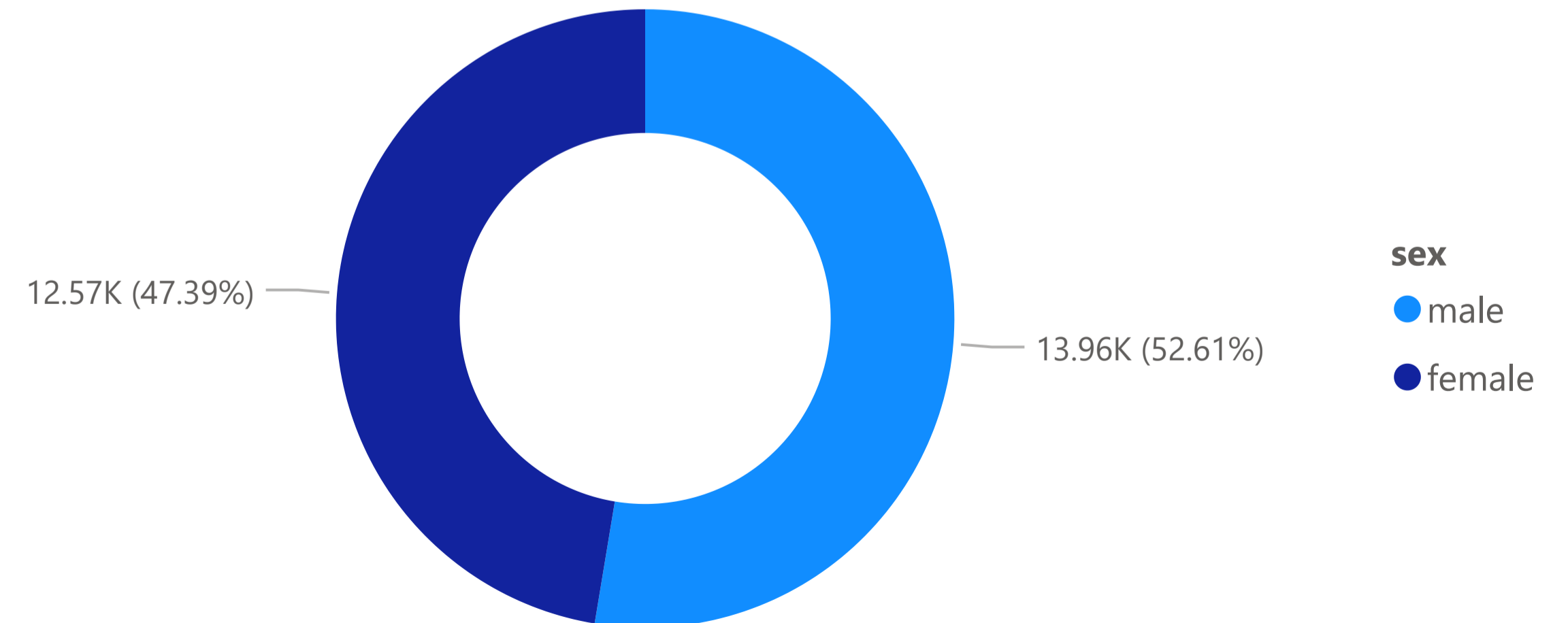
Sex vs Children: Average of children for male (1.12) was higher than female (1.07). female had 1.07 Average of children and male had 1.12.

Sex Vs Charges: Average of charges for male (13,956.75) was higher than female (12,569.58).

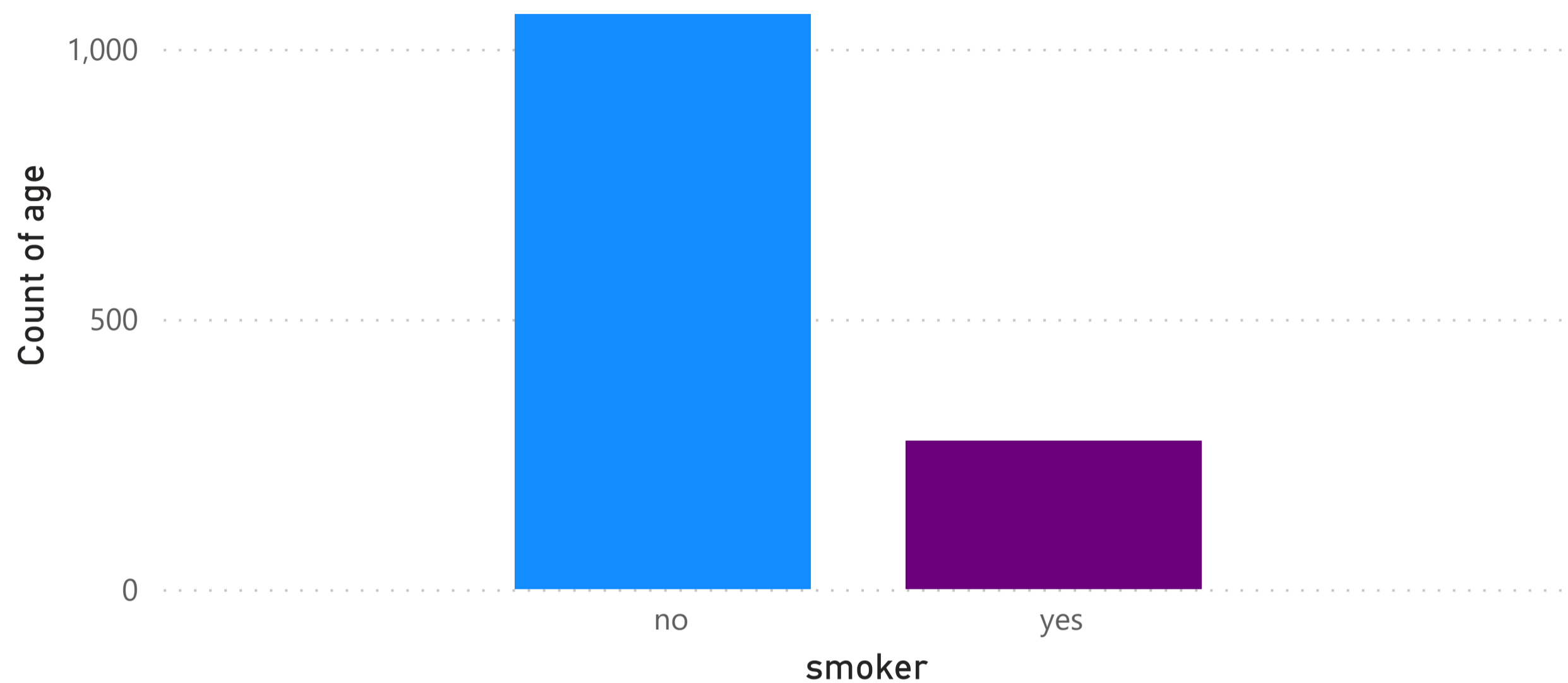
Average of children by sex



Average of charges by sex



Count of age by smoker

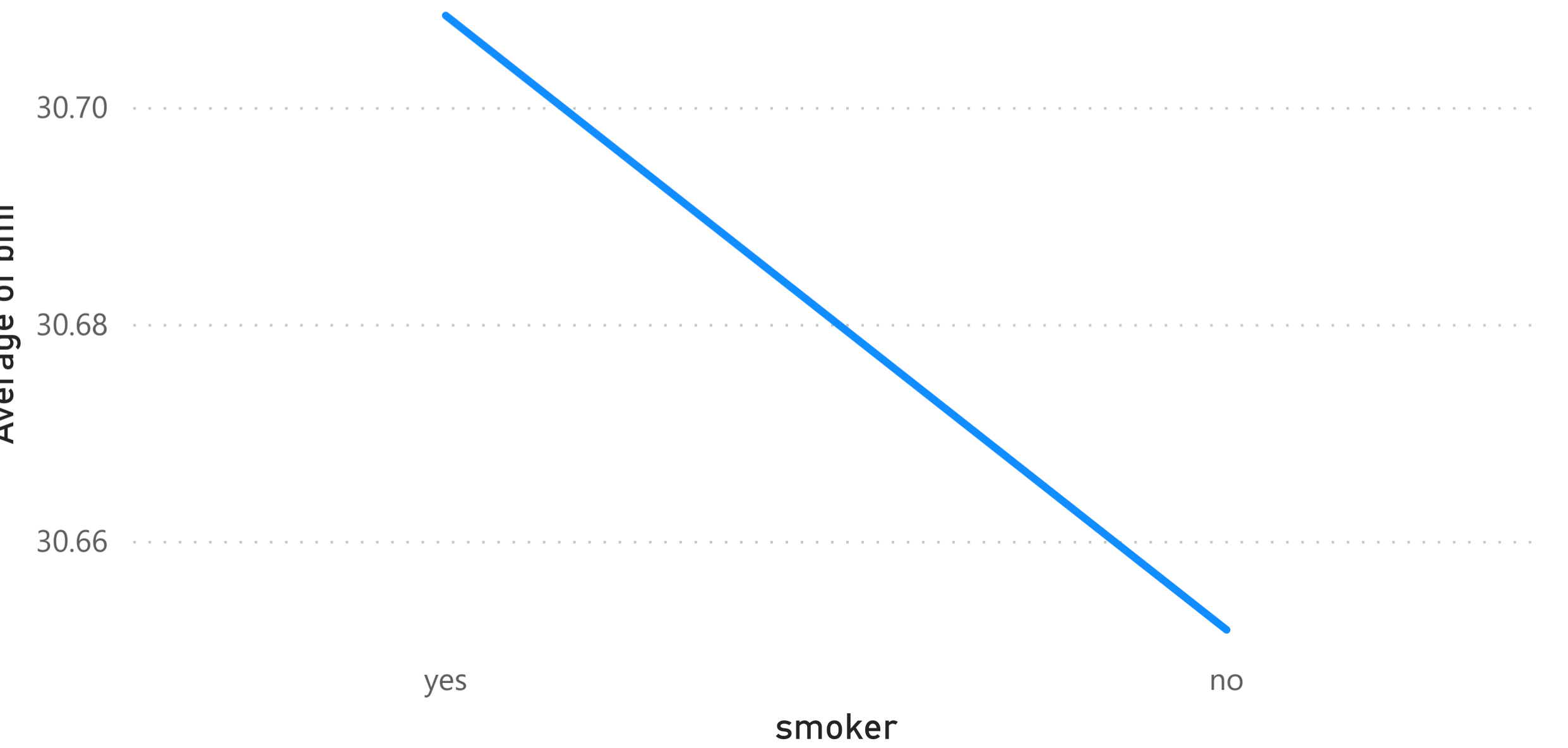


Smoker Vs Age: Count of age for no (1,064) was higher than yes (274).
 no accounted for 79.52% of Count of age.
 no had 1,064 Count of age and yes had 274.

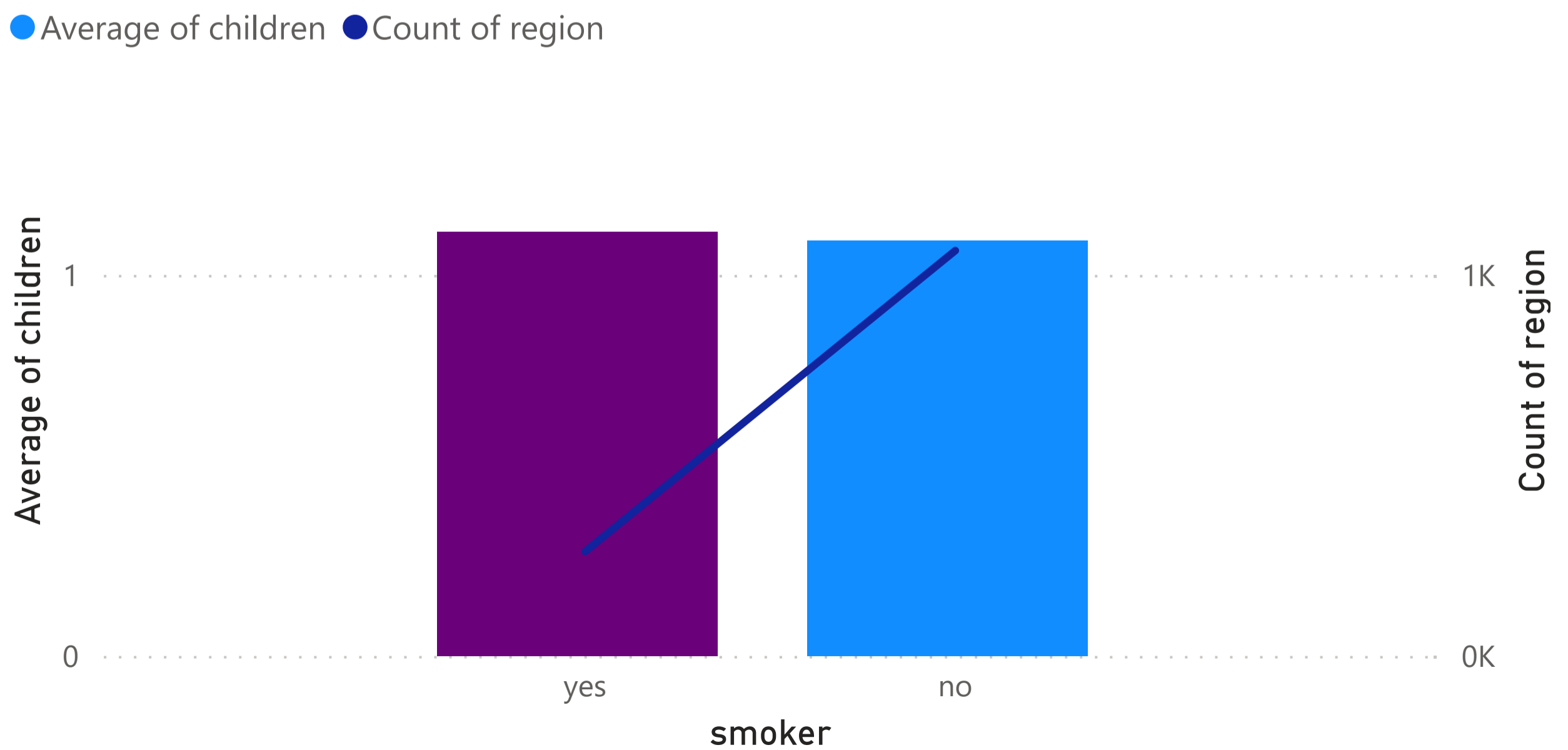
Smoker Vs BMI: Average of bmi for yes (30.71) was higher than no (30.65).
 .no had 30.65 Average of bmi and yes had 30.71.

Smoker Vs Children & Region: Average of children and total Count of region are negatively correlated with each other.
 no accounted for 79.52% of Count of region.
 Count of region and Average of children diverged the most when the smoker was no, when Count of region were 1,062.91 higher than Average of children.

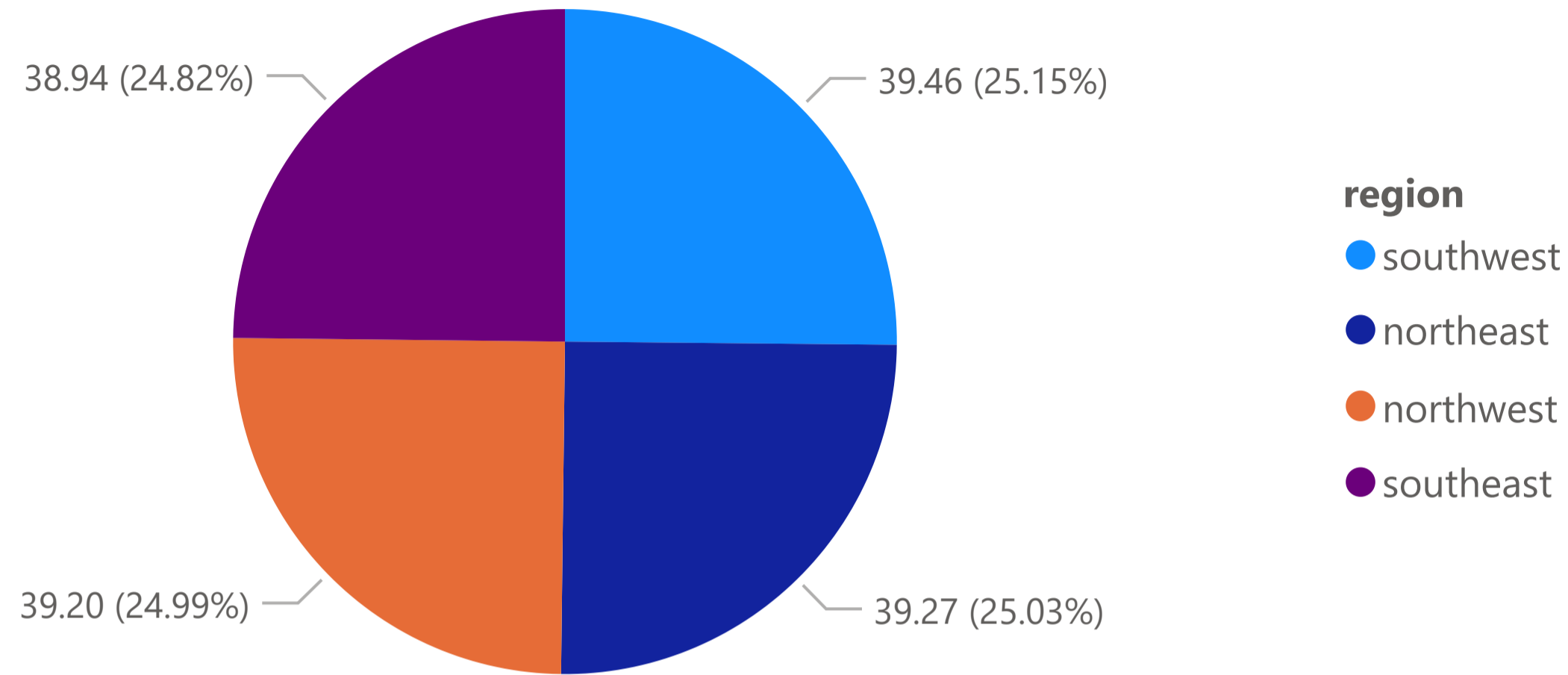
Average of bmi by smoker



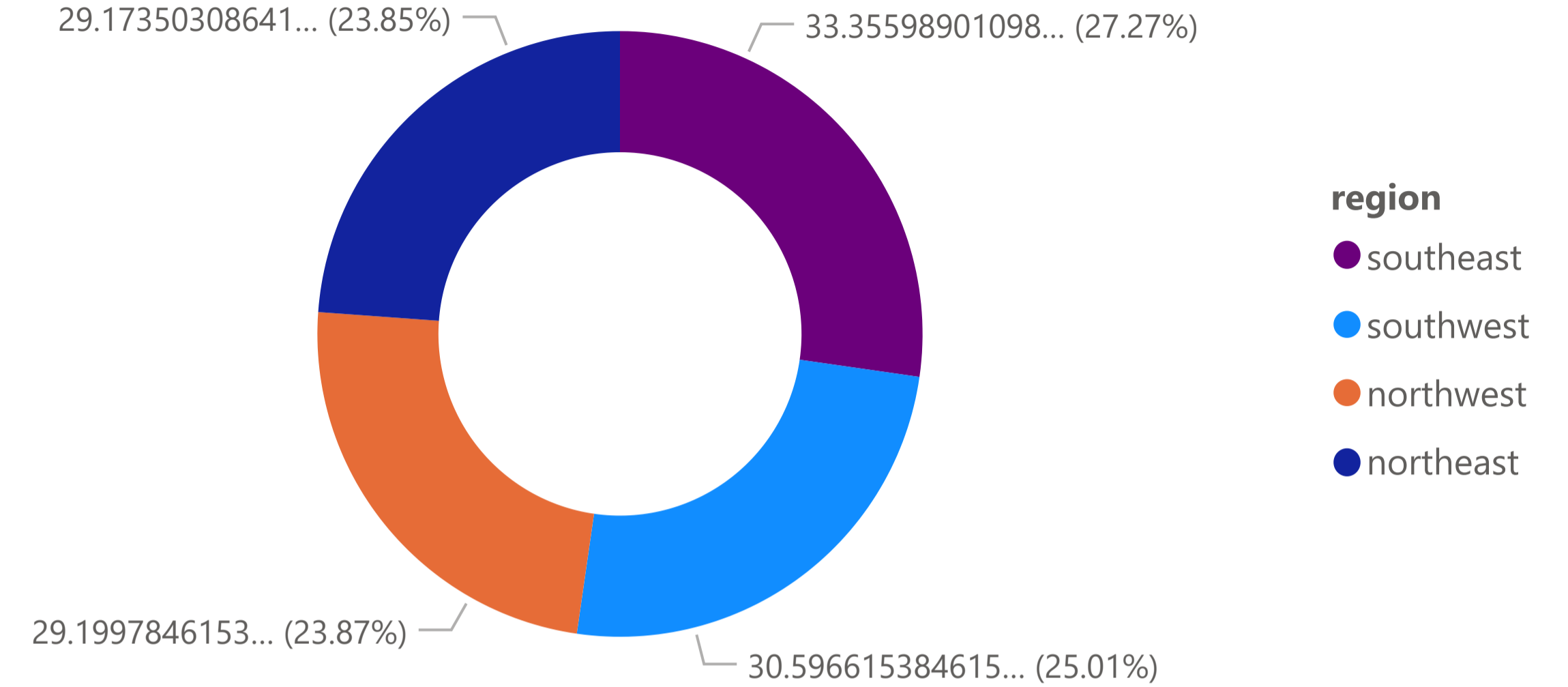
Average of children and Count of region by smoker



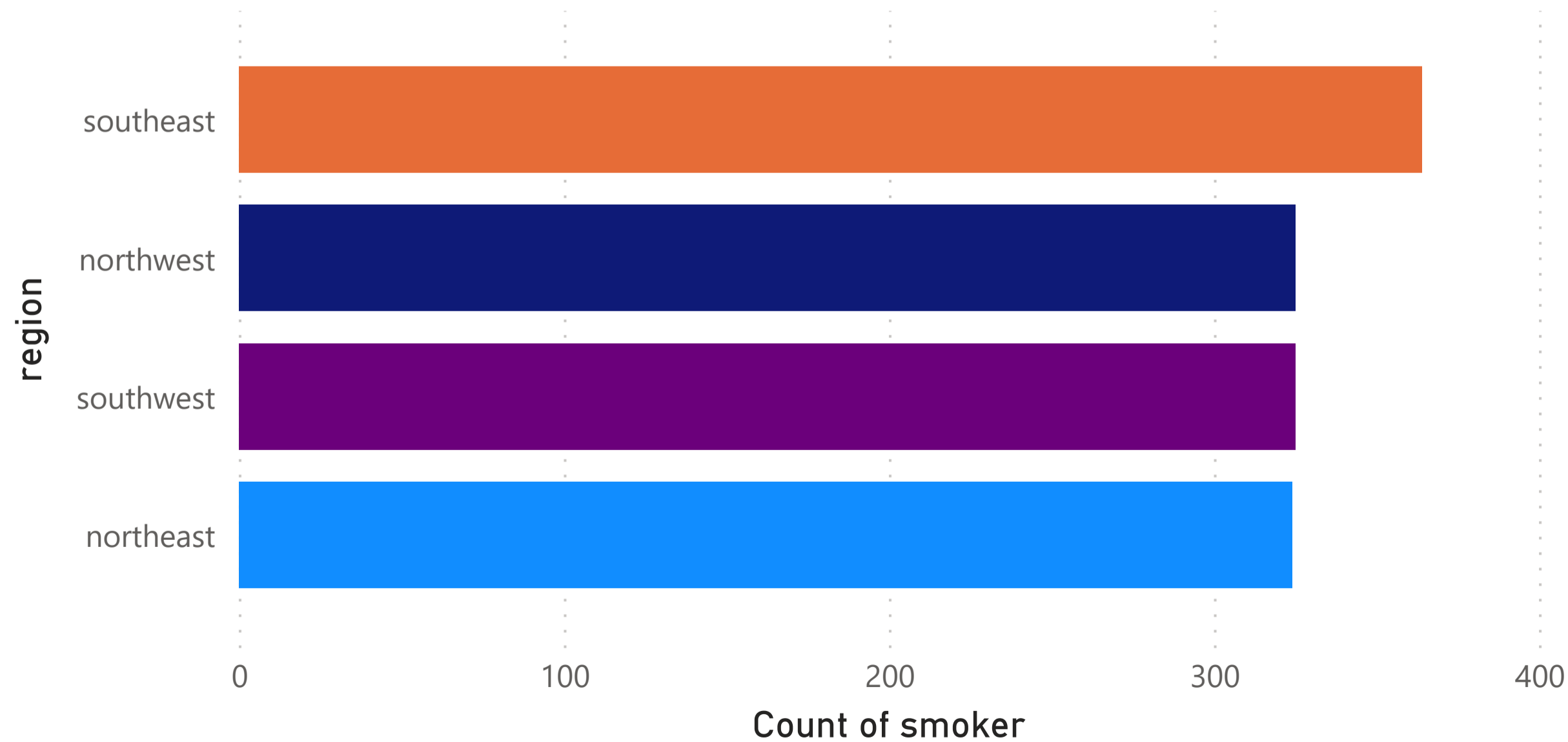
Average of age by region



Average of bmi by region



Count of smoker by region



Region Vs Avg Age: southwest had the highest Average of age at 39.46, followed by northeast, northwest, and southeast.

Region Vs Smoker: At 364, southeast had the highest Count of smoker and was 12.35% higher than northeast, which had the lowest Count of smoker at 324.

southeast had the highest Count of smoker at 364 and northeast had the lowest at 324.

southeast accounted for 27.20% of Count of smoker.

Across all 4 region, Count of smoker ranged from 324 to 364.

Region Vs BMI: southeast had the highest Average of bmi at 33.36, followed by southwest, northwest, and northeast.

Avg Charges Vs Smoking: Average of charges for yes (32,050.23) was higher than no (8,434.27).

no had 8,434.27 Average of charges and yes had 32,050.23.

Avg Charges Vs BMI: We see a low average correlation of BMI and Charges, with a few outliers towards the upper realm, which are like paired with Smoking.

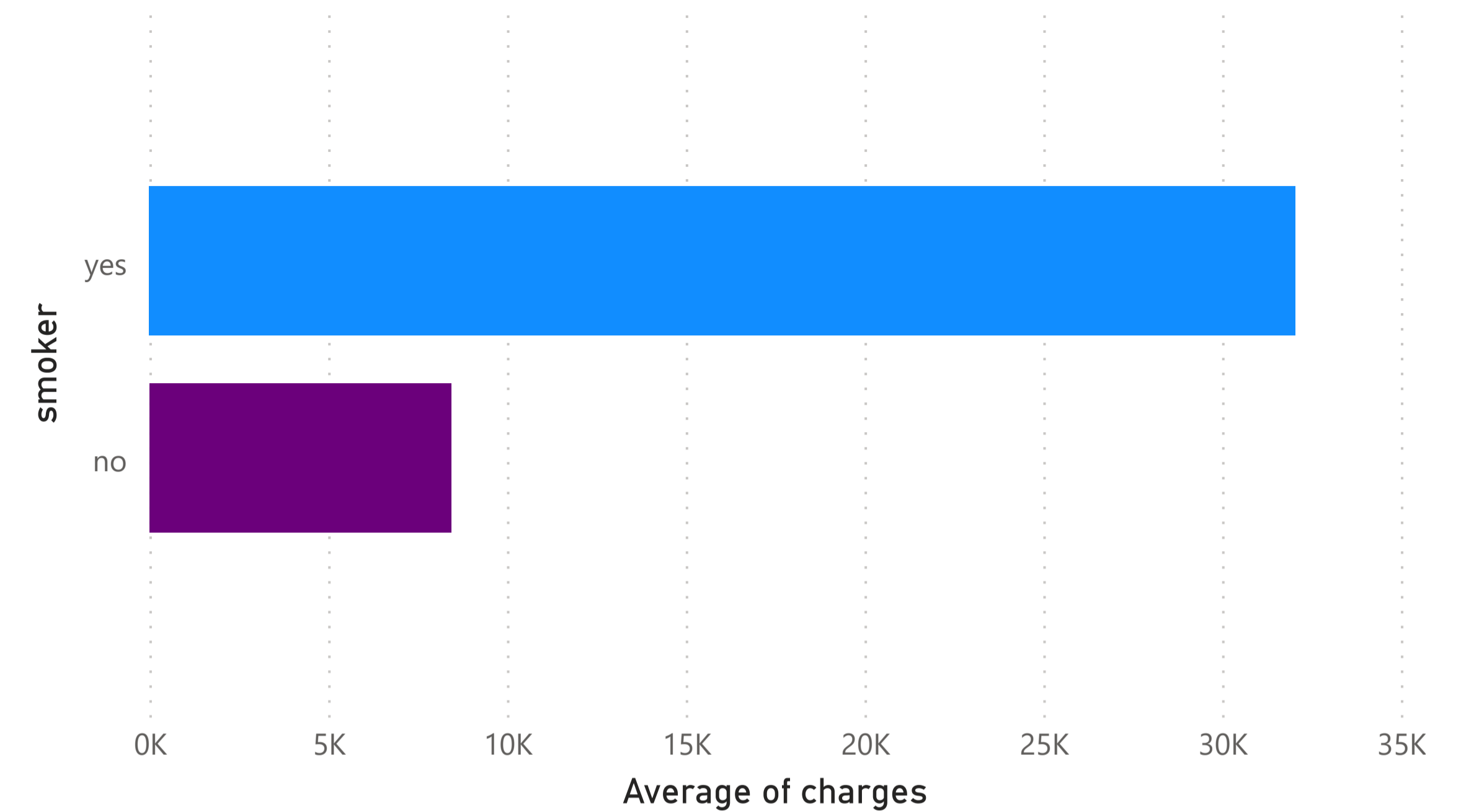
Avg Charges Vs Age: At 23,275.53, 64 had the highest Average of charges and was 392.03% higher than 21, which had the lowest Average of charges at 4,730.46.

Across all 47 age, Average of charges ranged from 4,730.46 to 23,275.53.

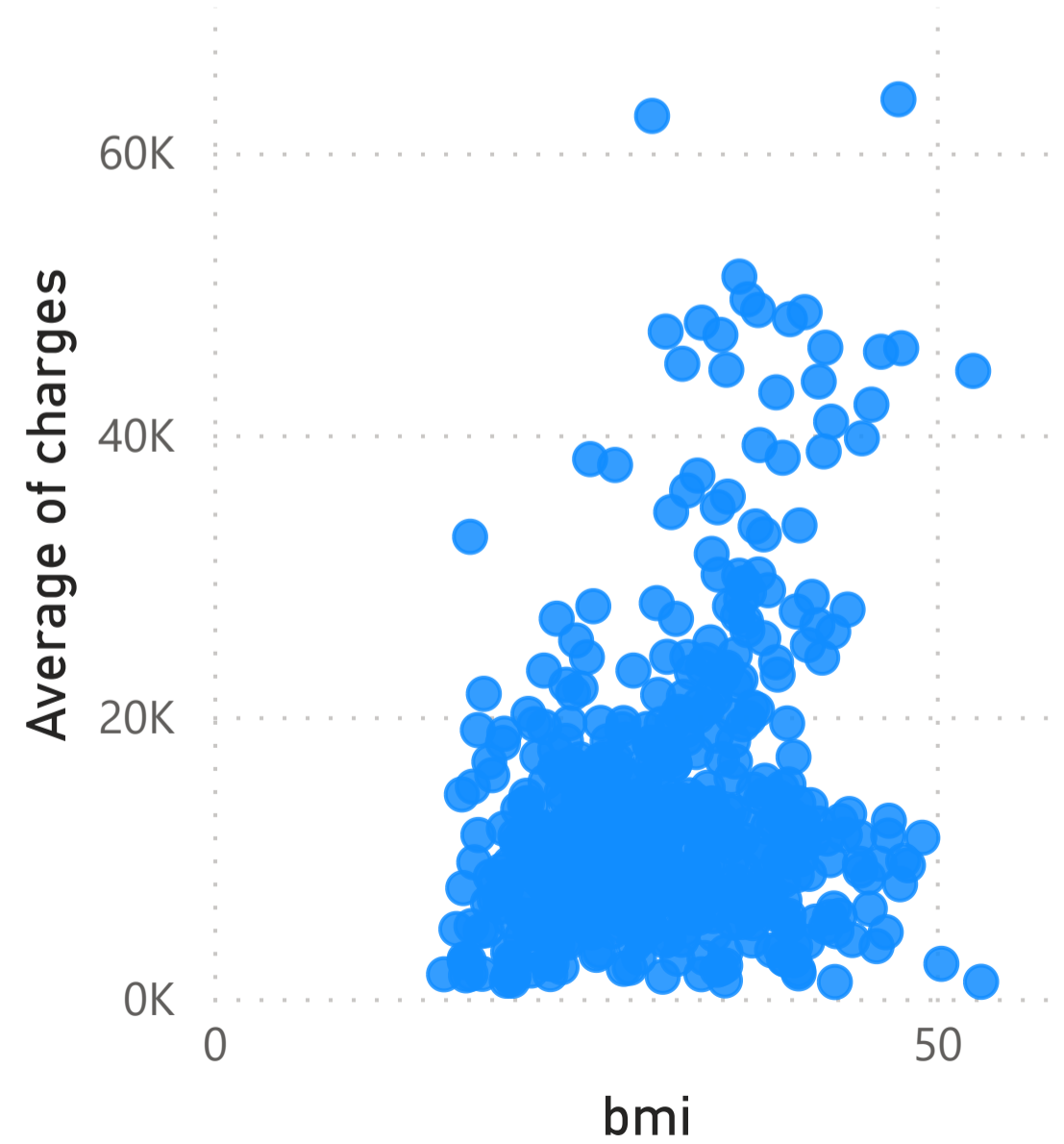
Avg Charges Vs Children: At 15,355.32, 3 had the highest Average of charges and was 74.77% higher than 5, which had the lowest Average of charges at 8,786.04.

Across all 6 children, Average of charges ranged from 8,786.04 to 15,355.32.

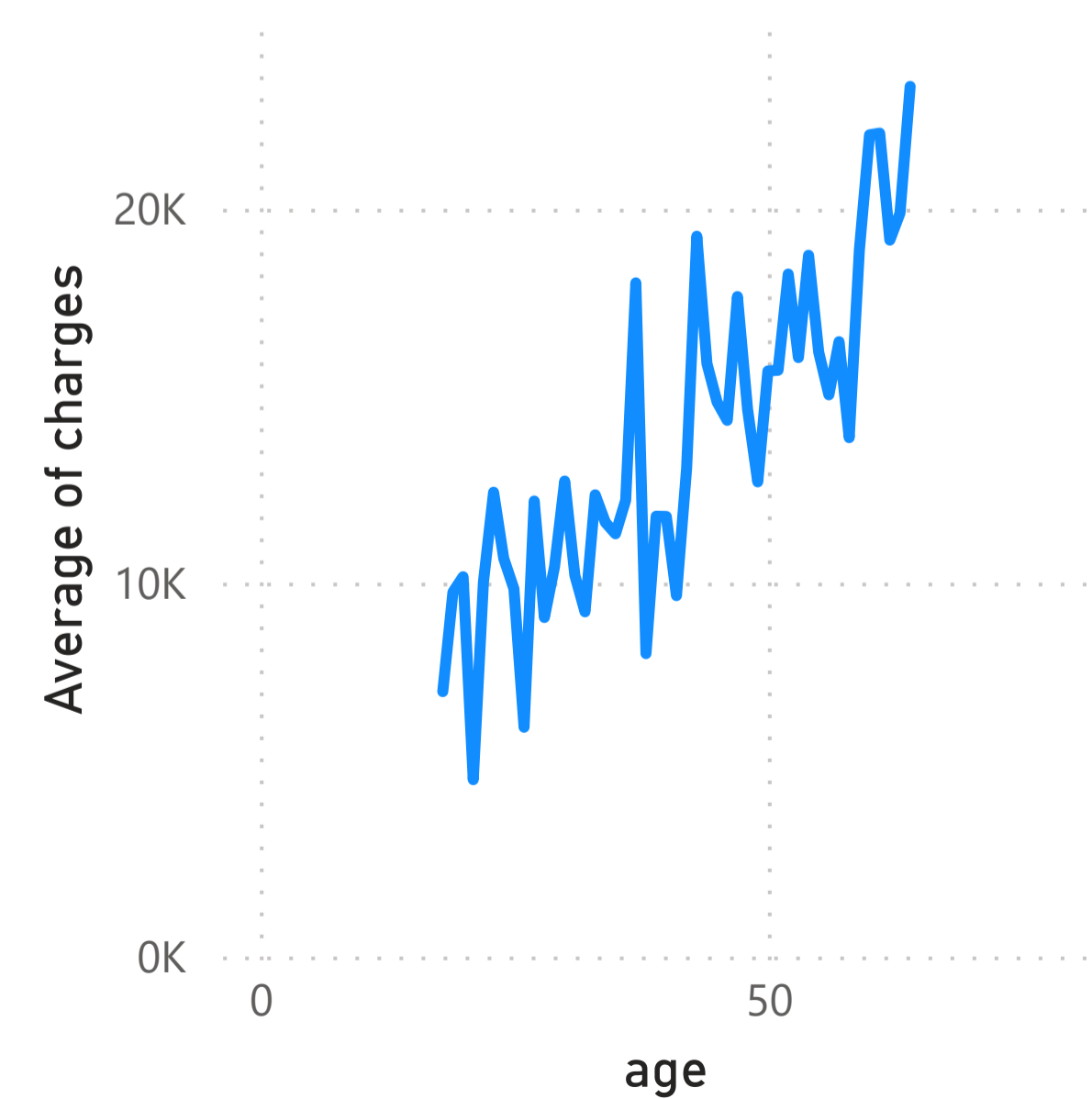
Average of charges by smoker



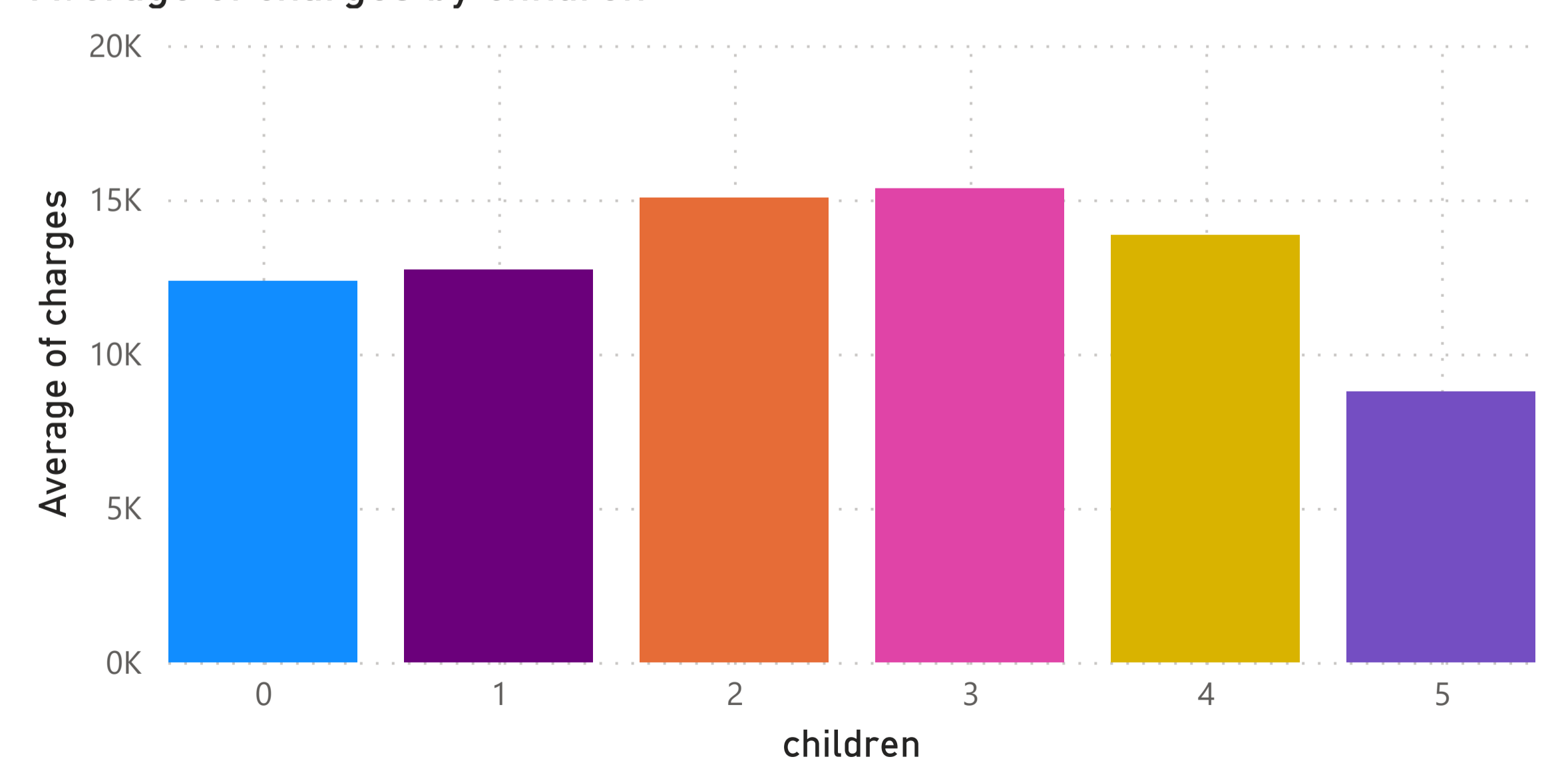
Average of charges by bmi



Average of charges by age



Average of charges by children

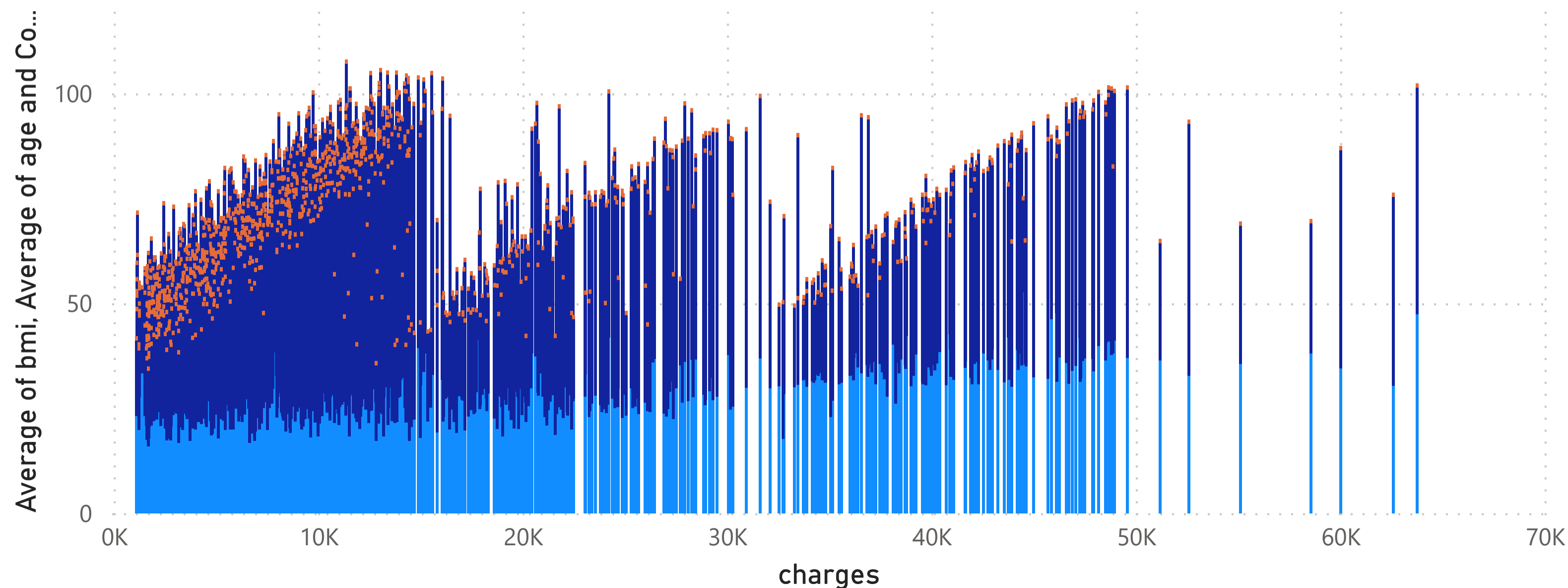


Regression Statistics								
Multiple R	0.864567614							
R Square	0.747477159							
Adjusted R Square	0.746909266							
Standard Error	6092.319464							
Observations	1338							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	1.46561E+11	48853667351	1316.230148	0			
Residual	1334	49513219514	37116356.46					
Total	1337	1.96074E+11						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-11676.83043	937.568702	-12.4543731	9.20815E-34	-13516.10009	-9837.560756	-13516.10009	-9837.560756
Age	259.5474916	11.93417597	21.7482541	5.24116E-90	236.1356948	282.9592883	236.1356948	282.9592883
BMI	322.6151328	27.48741481	11.73683066	2.41856E-30	268.6918649	376.5384008	268.6918649	376.5384008
Smoker	23823.6845	412.8666843	57.70309255	0	23013.7458	24633.62319	23013.7458	24633.62319

- R Square: The coefficient of determination (R^2) is 0.7475. It indicates that approximately 74.75% of the variability in the charges can be explained by the linear regression model with the given predictors. This suggests that the model fits the data reasonably well.
- Standard Error: The standard error is 6092.3195. It represents the average amount of variation or dispersion in the observed charges that is not accounted for by the regression model. A lower standard error indicates a better fit of the model to the data.
- Coefficients: The coefficients represent the estimated effects of the predictor variables on the charges.
 - Intercept: The intercept is -11676.8304. It represents the estimated charges when all the predictor variables are zero.
 - Age: The coefficient for age is 259.5475. It suggests that, holding other predictors constant, each unit increase in age is associated with an estimated increase of \$259.55 in charges.
 - BMI: The coefficient for BMI is 322.6151. It indicates that, holding other predictors constant, each unit increase in BMI is associated with an estimated increase of \$322.62 in charges.
 - Smoker: The coefficient for being a smoker is 23823.6845. It suggests that, holding other predictors constant, being a smoker is associated with an estimated increase of \$23,823.68 in charges compared to non-smokers.
- Significance: The p-values for the coefficients indicate the statistical significance of the estimated effects. In this case, all predictor variables (Age, BMI, and Smoker) have p-values close to zero, indicating that their effects on charges are statistically significant.
- Confidence Intervals: The lower 95% and upper 95% values represent the confidence intervals for the estimated coefficients. They provide a range within which the true population values of the coefficients are likely to fall with a certain level of confidence.

Average of bmi, Average of age and Count of smoker by charges

● Average of bmi ● Average of age ● Count of smoker

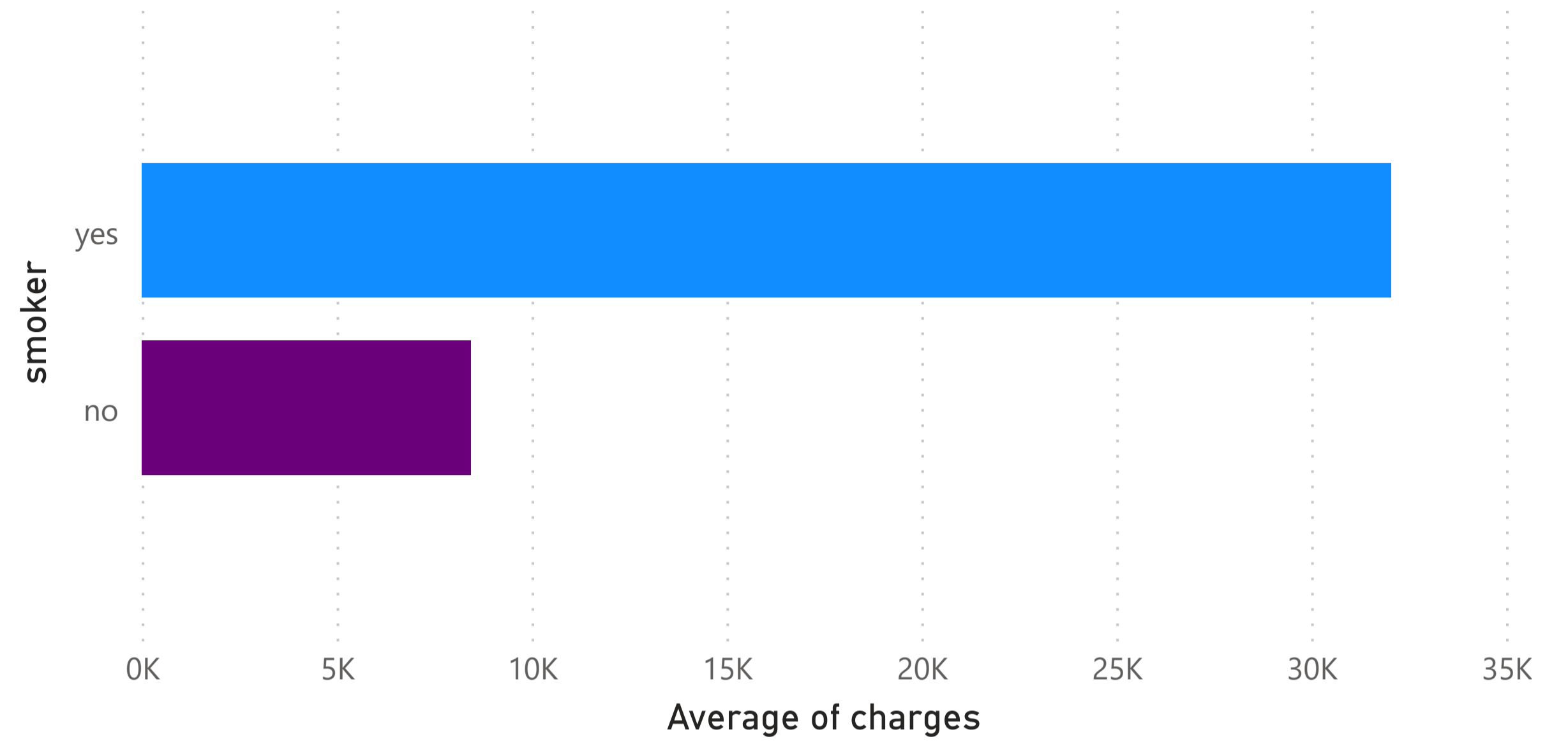


Overall, the regression analysis suggests that age, BMI, and smoking status significantly influence medical insurance charges. Age and BMI have positive effects, indicating that as individuals get older or have a higher BMI, their charges tend to increase. Being a smoker has the most significant effect on charges, with smokers having substantially higher charges compared to non-smokers.

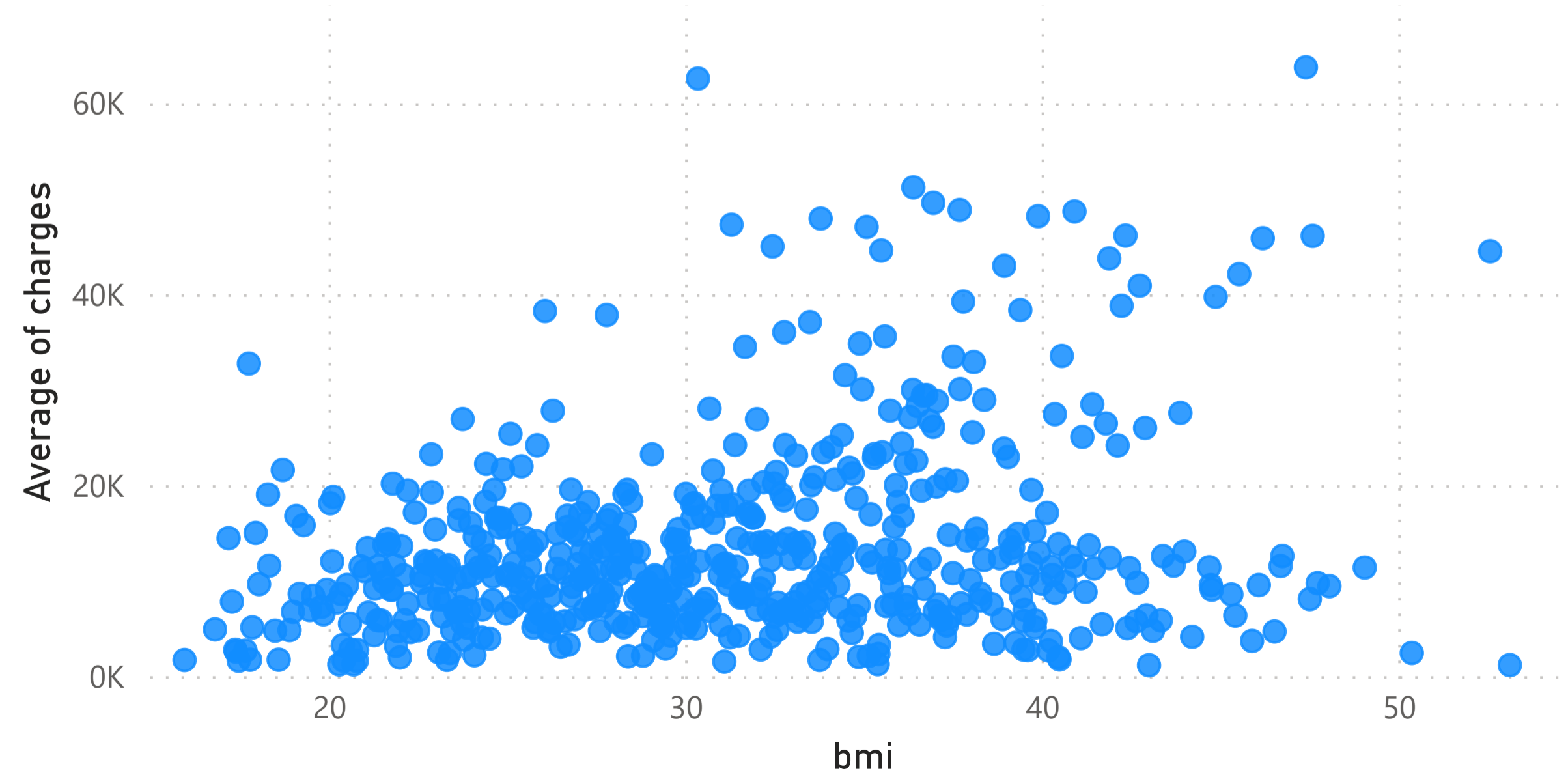
Summary: Albeit we have a large correlation with smoking; followed by BMI and Age, the data set is too limited to create an accurate prediction of potential pricing for insurance rates.

Parts of the process have consisted of a regression analysis which averaged 58% accuracy because of the limited data from this set.

Average of charges by smoker



Average of charges by bmi



Average of charges by age

