	<pre>import csv import pandas as pd import seaborn as sns import matplotlib.pyplot as plt</pre>
In [10]:	df = pd.read_csv("C:\Users\natha\Desktop\Coding_Study\Portfolio\\insurance.csv")    age   sex   bmi   children   smoker   region   charges     1   18   male   33.77   1   no   southeast   1725.55230     2   28   male   33.00   3   no   southeast   4449.46200
	3         male         2.705         0         no         northwest         21984.47061           4         32         male         2.880         0         no         northwest         386.85520           1335         7         male         3.970         3         no         northwest         1600.54830           1336         18         female         3.890         0         no         northwest         205.98080           1336         21         female         3.880         0         no         southeast         1629.83350           1336         21         female         2.880         0         no         southeast         1629.83350
In [29]:	1337 61 female 29.070 0 yes northwest 29141.36030  1338 rows × 7 columns  sns.set_style("white") sns.set_palette("pastel") age = df["age"] sns.despine() plt.show()
	200   175   150   125   150   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100   100
In [23]:	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
	sns.displot(sex) plt.show()  700 600 500 400
	8 300 200 100 100 1 female male
	children = df["children"] sns.displot(children) plt.show()  600  500
	400
	<pre>x = df['age'] y = df['children'] plt.scatter(x, y, alpha=0.5) sns.regplot(data=df, x="age", y="children")</pre>
	plt.show()  5
	1 0 20 30 40 50 60 age
	350 300 250 150
In [27]:	100 50 southwest southeast northwest region northeast smoker = df["smoker"]
	sns.displot(smoker) plt.show()  800
	To       400       200       0
	sns.violinplot(data=df, x="region", y="age") plt.show()  70  60  50  70  60  70  60  70  60  70  60  70  60  70  60  70  60  70  60  70  60  70  60  70  60  6
	region_mean_age = df.groupby('region').age.mean()
Out[32]: In [36]:	region mean_age.head()  region northeast
In [37]: Out[37]:	region northeast 13406.384516 northeast 12417.575374 southeast 12346.937377 Name: charges, dtype: float64  region_mean_children = df.groupby('region').children.mean() region_mean_children.head()  region northeast 1.046296 northwest 1.147692
In [54]:	<pre>southeast 1.049451 southwest 1.141538 Name: children_40 = [] for i in df['age']:     if i == 40:         children_40.append(df['children']) print(len(children_40))</pre>
	new_df = df.loc[df['age'] == 40] print(new_df)   age
	235         40         female         22.220         2         yes         southeast         19444.26580           243         40         male         35.300         3         no         southeast         7196.86700           266         40         female         28.120         1         yes         southeast         17179.52200           325         40         female         28.120         1         yes         northeast         6600.20595           422         40         male         32.775         1         yes         northeast         5910.94400           459         40         female         22.705         2         no         northeast         7173.35995           667         40         female         23.370         3         no         northeast         7173.35995           667         40         female         23.775         2         yes         northeast         7173.35995           67         40         female         23.775         2         yes         northeast         7173.35995
In [66]:	669
In [68]:	<pre>children_40.extend(new_df['children'].tolist()) print(children_40)  [3, 1, 0, 1, 1, 4, 2, 3, 1, 1, 1, 1, 0, 3, 2, 3, 2, 1, 0, 2, 1, 2, 1, 1, 0, 2, 4]  ave_children = sum(children_40)/len(children_40) print(ave_children)  1.5925925925925926  bmi_list =[] bmi_list.extend(df['bmi'].tolist())</pre>
	<pre>ave_bmi = sum(bmi_list)/len(bmi_list) print(ave_bmi)  no_of_people_above_ave = 0 for i in bmi_list:     if i &gt; ave_bmi:         no_of_people_above_ave += 1 print(no_of_people_above_ave) print(round((no_of_people_above_ave/len(bmi_list)*100),2))</pre> 30.663396860986538
In [155	<pre>466 48.28  age = [] sex = [] children = [] smoker = [] region = [] charges = []  age.extend(df['age'].tolist()) sex.extend(df['sex'].tolist())</pre>
In [153	<pre>children.extend(df['children'].tolist()) smoker.extend(df['smoker'].tolist()) region.extend(df['region'].tolist()) charges.extend(df['charges'].tolist())  combined_list = list(zip(age, sex, bmi_list, children, smoker, region, charges)) #print(combined_list)  Dict_Com = {} for i in range(0, len(combined_list)):</pre>
In [154 In [94]:	Dict_Com[("Customer " + str(i))] = combined_list[i]  #print(Dict_Com)  print(Dict_Com["Customer 0"])  (19, 'female', 27.9, 0, 'yes', 'southwest', 16884.924)  print(Dict_Com["Customer 36"])  (62, 'female', 32.965, 3, 'no', 'northwest', 15612.19335)
In [96]:	<pre>print(Dict_Com["Customer 44"])  (38, 'male', 37.05, 1, 'no', 'northeast', 6079.6715)  x = df['age'] y = df['charges']  plt.scatter(x, y, alpha=0.5) sns.regplot(data=df, x="age", y="charges")  plt.show()</pre>
	60000
	20 30 40 50 60  x = df['bmi'] y = df['charges'] plt.scatter(x, y, alpha=0.5) sns.regplot(data=df, x="bmi", y="charges") plt.show()
	6000 5000 4000 6 3000 2000
	<pre>bmi_charge_list = list(zip(bmi_list, charges)) bmi_large_than_ave_charge = [] charge_for_bmi_above_ave = [] for item in bmi_charge_list:     if item[0] &gt; ave_bmi:</pre>
In [122	<pre>bmi_large_than_ave_charge.append(item[0])</pre>
	#sns.regplot(x="bmi", y="charges") plt.show()  60000 40000
In [124	2000 10000 30 35 40 45 50 new_df_1 = df.loc[df['charges'] > 30000] print(new_df_1) print(len(new_df_1))
	age         sex         bmi         children smoker         region         charges           14         27         male         42.130         0         yes         southwest         39611.75770           19         30         male         31.530         0         yes         southwest         36837.46700           23         34         female         31.920         1         yes         southwest         37701.87680           30         22         male         35.600         2         yes         southwest         35711.0000           30         22         male         35.600         yes         southwest         3585.57600           1301         62         male         27.800         yes         southwest         3782.72420           1303         43         male         27.800         yes         southwest         33900.65300           1313         19         female         34.700         yes         southwest         36397.57600
In [125	1323 42 female 40.370 2 yes southeast 43896.37630  [162 rows x 7 columns] 162  smoker = new_df_1["smoker"] sns.displot(smoker) plt.show()
	120 100 100 80 40
	<pre>pes</pre>
In [129	number_of_smoker +=1  print(number_of_smoker)  152  perc_number_of_smoker = (number_of_smoker/len(new_df_1))*100 print(perc_number_of_smoker)  93.82716049382715  number_high_bmi = 0
	<pre>for item in new_df_1["bmi"]:     if item &gt;ave_bmi:         number_high_bmi +=1  print(number_high_bmi) perc_number_high_bmi = (number_high_bmi/len(new_df_1))*100 print(perc_number_high_bmi)  140 86.41975308641975</pre>
	sex = new_df_1["sex"] sns.displot(sex) plt.show()  100 80
	region = new_df_1["region"] sns.displot(region) plt.show()  00 50
	40
	print(sum(new_df_1['children']))/len(new_df_1['children'])) print(sum(df['children']))/len(df['children'])) print(sum(df['children']))/len(df['children'])) print(len(df['children']))
In [141	1.166666666666667 162 1.0949177877429 1338  age_1 = new_df_1["age"] sns.displot(age_1) sns.desplot() plt.show()
	25 20 Tung 15 15 15 16 16 16 16 16 16 16 16 16 16 16 16 16
	smoker_list = [] sex_list = []
	<pre>smoker_list.extend(new_df_1['smoker'].tolist()) sex_list.extend(new_df_1['sex'].tolist())  com_sex_smoker = list(zip(smoker_list, sex_list)) #print(com_sex_smoker)  count = 0 for item in com_sex_smoker:     if item[0] == 'yes' and item[1] == 'male':         count += 1</pre>
	<pre>print(count)  count = 0 for item in com_sex_smoker:     if item[0] == 'yes' and item[1] == 'male':</pre>
In [147	<pre>print(count/number_of_smoker)  0.6447368421052632  count_female = 0 for item in com_sex_smoker:     if item[0] == 'yes' and item[1] == 'female':         count_female += 1  print(count_female)  54</pre>
In [149	<pre>print(count_female/number_of_smoker) 0.35526315789473684</pre>
In [151	<pre>female = 0 for item in sex_list:     if item == 'female':         female += 1  print(female)  59  new_df_2 = new_df_1.loc[df['smoker'] == 'no'] print(new_df_2)</pre>
	print (new_df_2)           ge         sex         bmi         children smoker         region         charges           62         64         male         24.70         1         no         northeest         30166.61817           115         60         male         28.595         0         no         northeest         30259.9955           242         55         female         25.365         2         no         northeest         3160.13457           387         50         male         25.355         2         no         northeest         3162.01016           59         62         female         37.525         2         no         northeest         33471.97189           936         44         male         29.735         2         no         northeest         3108.66282           102         61         female         33.300         4         no         southeest         36580.28216           102         69         female         34.800         2         no         southeest         36590.6803
	1258 55 male 37.715 3 no northwest 30063.58055 10