

In [1]:

```
import pandas as pd
```

In [2]:

```
df = pd.read_csv('driverresponse.csv')
```

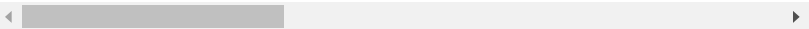
In [3]:

```
df.head()
```

Out[3]:

	index	sno	stateut	region	regionid	alcintake2014	overspeed2014	overtaking2014	lan
0	0	1	Andhra Pradesh	south	2	594	12747.0	507	
1	1	2	Arunachal Pradesh	northeast	5	11	16.0	0	
2	2	3	Assam	northeast	5	613	4596.0	129	
3	3	4	Bihar	north	1	1680	1496.0	278	
4	4	5	Chhattisgarh	centre	9	335	6720.0	188	

5 rows × 21 columns



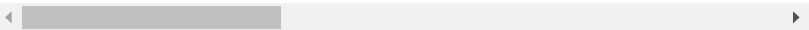
In [4]:

```
df.tail()
```

Out[4]:

	index	sno	stateut	region	regionid	alcintake2014	overspeed2014	overtaking2014	
31	31	32	D & N Haveli	west	4	1	21.0	0	
32	32	33	Daman & Diu	west	4	0	21.0	2	
33	33	34	Delhi	north	1	51	NaN	0	
34	34	35	Lakshadweep	southwest	8	0	1.0	0	
35	35	36	Puducherry	south	2	73	250.0	136	

5 rows × 21 columns



In [5]:

```
df.shape
```

Out[5]:

(36, 21)

In [6]:

```
df.columns
```

Out[6]:

```
Index(['index', 'sno', 'stateut', 'region', 'regionid', 'alcintake2014',  
      'overspeed2014', 'overtaking2014', 'lanejumping2014', 'wrongside2014',  
      'signalavoid2014', 'asleep2014', 'othercause2014', 'alcintake2016',  
      'overspeed2016', 'signalavoid2016', 'wrongside2016', 'lanejumping2016',  
      'overtaking2016', 'asleep2016', 'othercause2016'],  
      dtype='object')
```

In [7]:

```
df.duplicated().sum()
```

Out[7]:

```
0
```

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
index          0  
sno            0  
stateut        0  
region         0  
regionid       0  
alcintake2014  0  
overspeed2014  1  
overtaking2014 0  
lanejumping2014 0  
wrongside2014  0  
signalavoid2014 0  
asleep2014     0  
othercause2014 0  
alcintake2016  0  
overspeed2016  0  
signalavoid2016 0  
wrongside2016  0  
lanejumping2016 0  
overtaking2016 0  
asleep2016     0  
othercause2016 0  
dtype: int64
```

```
In [9]:
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   index                  36 non-null    int64
1   sno                    36 non-null    int64
2   stateut                36 non-null    object
3   region                 36 non-null    object
4   regionid               36 non-null    int64
5   alcintake2014          36 non-null    int64
6   overspeed2014          35 non-null    float64
7   overtaking2014         36 non-null    int64
8   lanejumping2014        36 non-null    int64
9   wrongside2014          36 non-null    int64
10  signalavoid2014        36 non-null    int64
11  asleep2014             36 non-null    int64
12  othercause2014         36 non-null    int64
13  alcintake2016           36 non-null    int64
14  overspeed2016           36 non-null    int64
15  signalavoid2016        36 non-null    int64
16  wrongside2016           36 non-null    int64
17  lanejumping2016         36 non-null    int64
18  overtaking2016          36 non-null    int64
19  asleep2016             36 non-null    int64
20  othercause2016          36 non-null    int64
dtypes: float64(1), int64(18), object(2)
memory usage: 6.0+ KB
```

```
In [10]:
df.describe()
```

Out[10]:

	index	sno	regionid	alcintake2014	overspeed2014	overtaking2014	lanejumpin
count	36.000000	36.000000	36.000000	36.000000	35.000000	36.000000	36.0
mean	17.500000	18.500000	4.000000	525.444444	5950.600000	312.833333	283.1
std	10.535654	10.535654	2.746426	767.133866	8677.731983	514.513279	462.5
min	0.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.0
25%	8.750000	9.750000	1.750000	11.750000	72.500000	1.500000	0.0
50%	17.500000	18.500000	4.000000	82.500000	2561.000000	75.500000	86.5
75%	26.250000	27.250000	5.000000	706.500000	6493.000000	350.000000	316.7
max	35.000000	36.000000	9.000000	3540.000000	29790.000000	2171.000000	1869.0

In [11]:

```
df.nunique()
```

Out[11]:

```
index          36
sno            36
stateut        36
region         9
regionid        9
alcintake2014  31
overspeed2014  33
overtaking2014 28
lanejumping2014 26
wrongside2014  29
signalavoid2014 19
asleep2014     19
othercause2014 30
alcintake2016  34
overspeed2016  35
signalavoid2016 22
wrongside2016  33
lanejumping2016 23
overtaking2016 33
asleep2016     24
othercause2016 32
dtype: int64
```

In [12]:

```
df = df.drop(['index', 'sno'], axis = 1)
```

In [13]:

```
df = df.dropna()
```

In [14]:

```
df.isnull().sum()
```

Out[14]:

```
stateut      0
region       0
regionid     0
alcintake2014 0
overspeed2014 0
overtaking2014 0
lanejumping2014 0
wrongside2014 0
signalavoid2014 0
asleep2014   0
othercause2014 0
alcintake2016 0
overspeed2016 0
signalavoid2016 0
wrongside2016 0
lanejumping2016 0
overtaking2016 0
asleep2016   0
othercause2016 0
dtype: int64
```

In [15]:

```
df['region'].unique()
```

Out[15]:

```
array(['south', 'northeast', 'north', 'centre', 'west', 'east',
       'northwest', 'southeast', 'southwest'], dtype=object)
```

In [16]:

```
df['region'].value_counts()
```

Out[16]:

```
north      8
northeast  6
south      5
centre     5
west       5
east       3
northwest  1
southeast  1
southwest  1
Name: region, dtype: int64
```

In [17]:

```
import matplotlib.pyplot as plt
import seaborn as sns
```

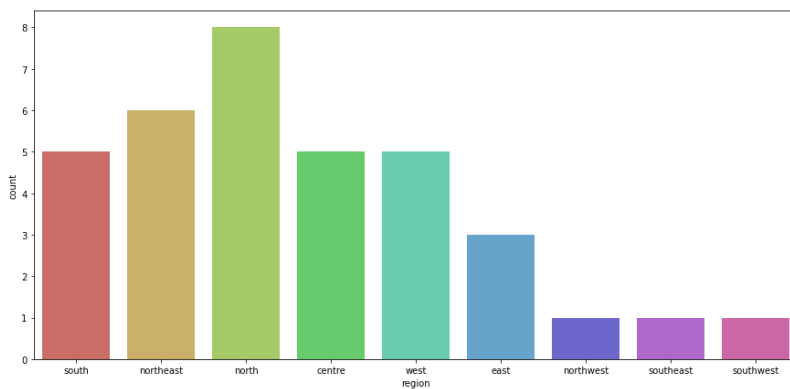
In [18]:

```
import warnings
warnings.filterwarnings('ignore')
```

In [19]:

```
plt.figure(figsize=[15,7],)
print('Countplot for Region')
sns.countplot(df['region'], data = df, palette = 'hls')
plt.xticks(rotation = 0)
plt.show()
```

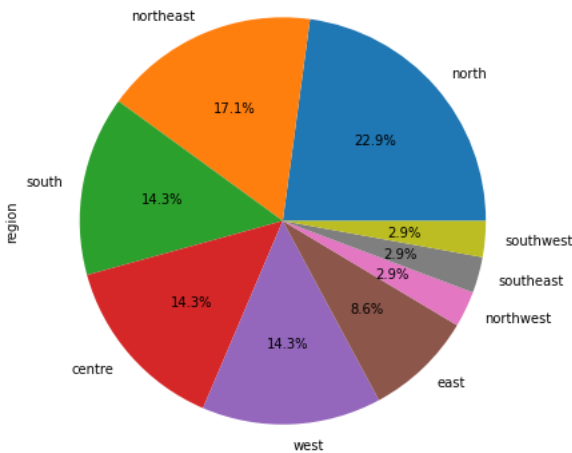
Countplot for Region



In [20]:

```
plt.figure(figsize=[15,7],)
print('Pieplot for Region')
df['region'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.xticks(rotation = 0)
plt.show()
```

Pieplot for Region



In [21]:

```
df['regionid'].unique()
```

Out[21]:

```
array([2, 5, 1, 9, 4, 3, 6, 7, 8], dtype=int64)
```

In [22]:

```
df['regionid'].value_counts()
```

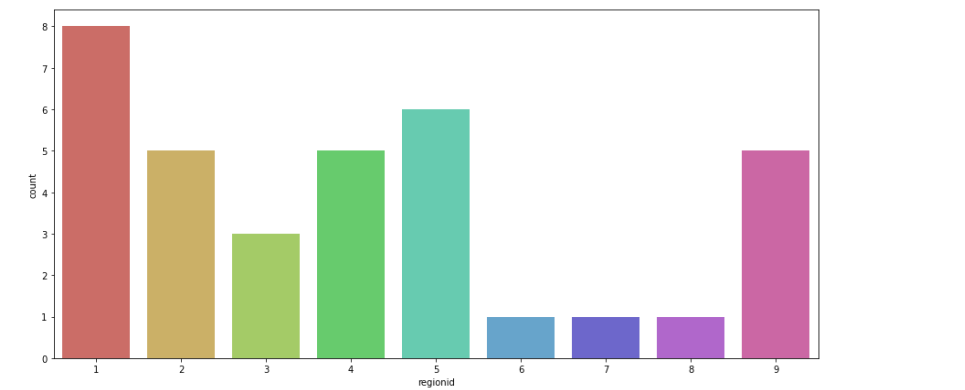
Out[22]:

```
1    8
5    6
2    5
9    5
4    5
3    3
6    1
7    1
8    1
Name: regionid, dtype: int64
```

In [23]:

```
plt.figure(figsize=[15,7],)
print('Countplot for RegionID')
sns.countplot(df['regionid'], data = df, palette = 'hls')
plt.xticks(rotation = 0)
plt.show()
```

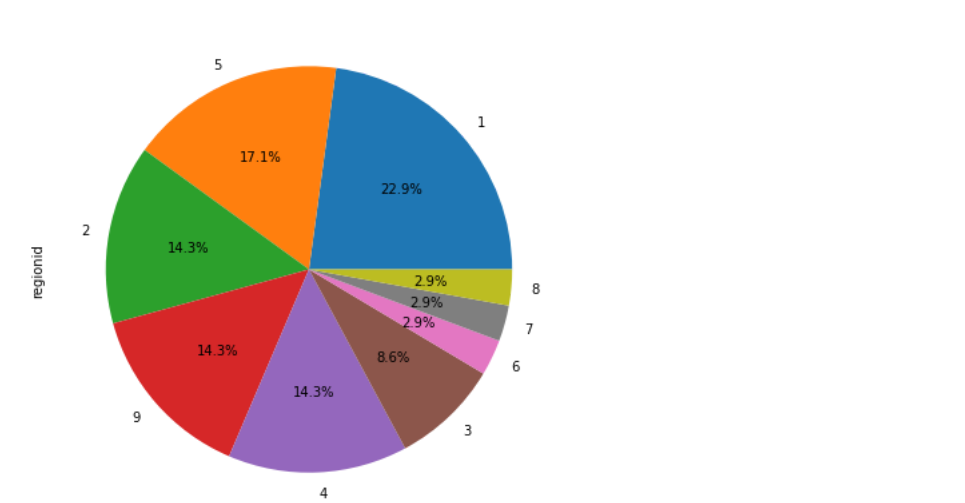
Countplot for RegionID



In [24]:

```
plt.figure(figsize=[15,7],)
print('Pieplot for RegionID')
df['regionid'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.xticks(rotation = 0)
plt.show()
```

Pieplot for RegionID

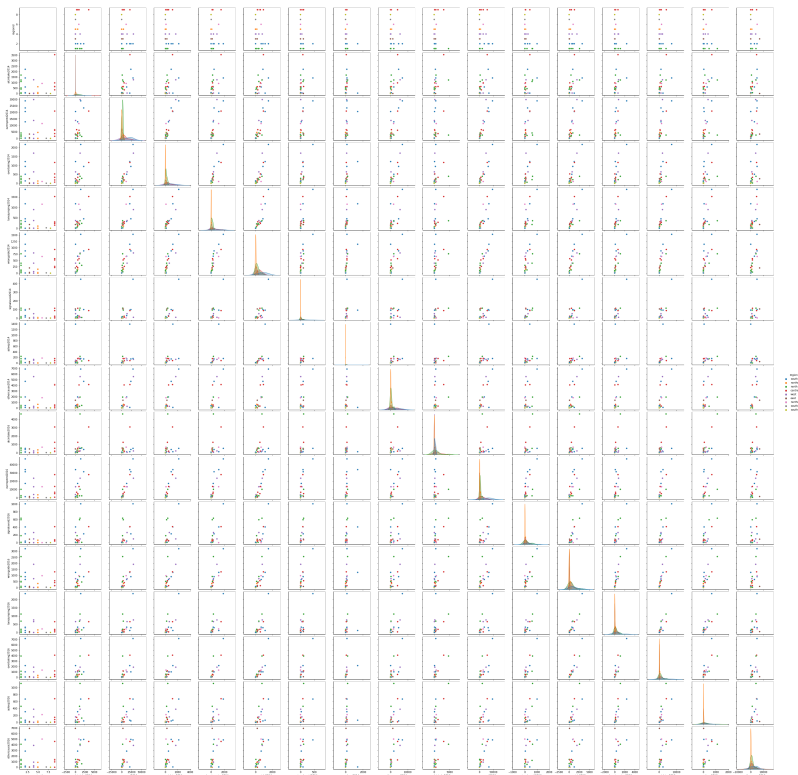


In [25]:

```
plt.figure(figsize=[15,7],)
print('Pairplot')
sns.pairplot(df, hue='region')
plt.xticks(rotation = 0)
plt.show()
```

Pairplot

<Figure size 1080x504 with 0 Axes>

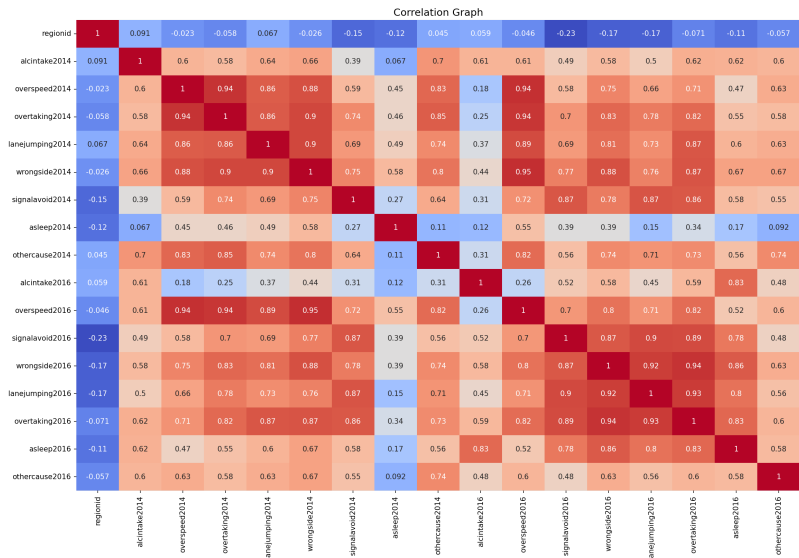


In [26]:

```
import numpy as np
```

In [27]:

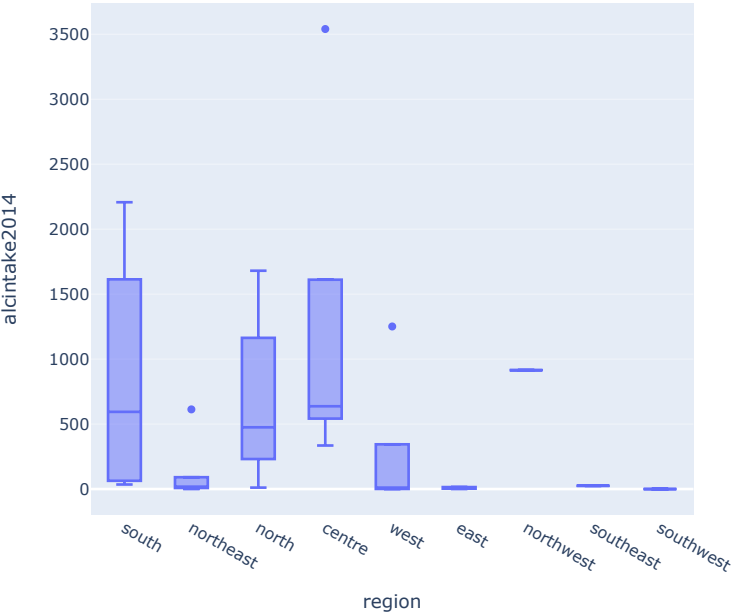
```
plt.figure(figsize = [18,12] , dpi = 150)
plt.title("Correlation Graph" , fontsize = 14)
matrix = np.triu(df.corr())
sns.heatmap(df.corr(), annot = True, cmap = 'coolwarm', cbar = False)
plt.show ()
```



In [28]:

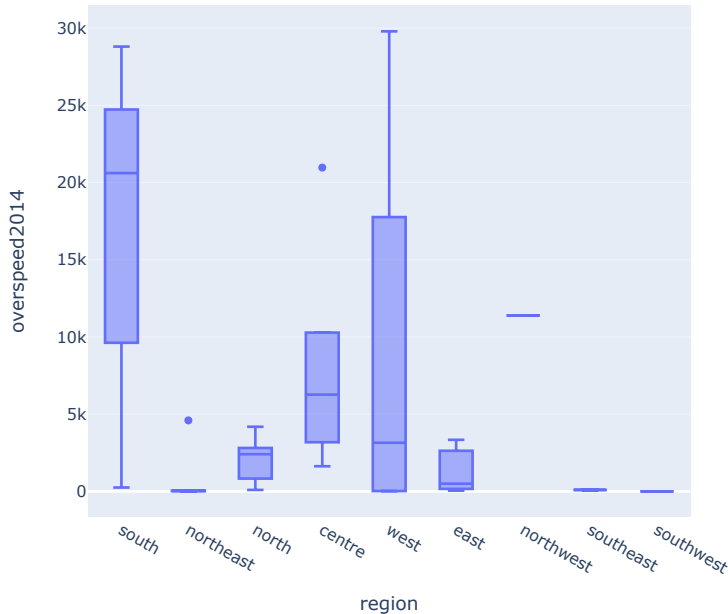
```
import plotly.express as px
```

```
In [29]:  
fig = px.box(df, x='region', y='alcintake2014')  
fig.show()
```

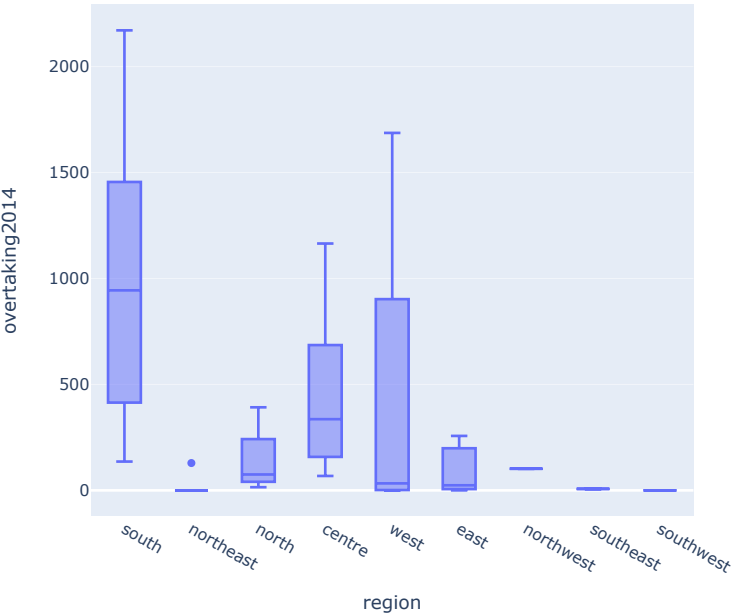


In [30]:

```
fig = px.box(df, x='region', y='overspeed2014')
fig.show()
```

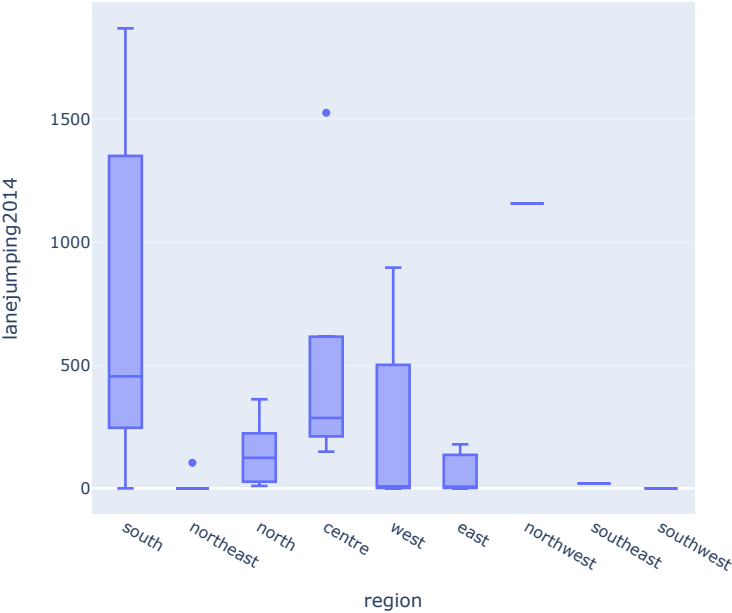


```
In [31]:  
fig = px.box(df, x='region', y='overtaking2014')  
fig.show()
```

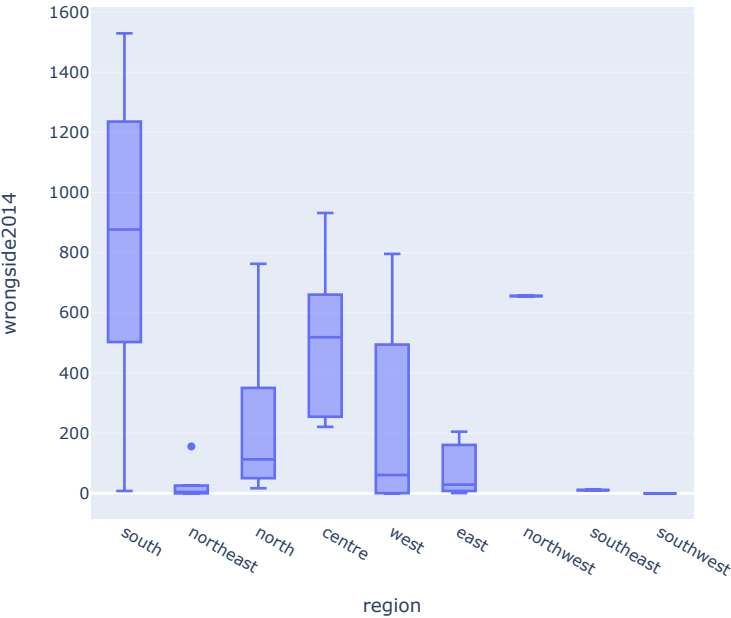


In [32]:

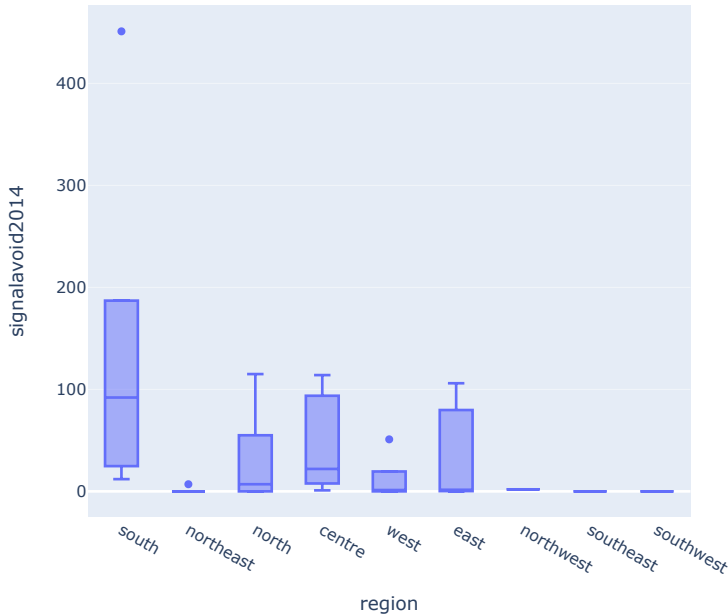
```
fig = px.box(df, x='region', y='lanejumping2014')
fig.show()
```



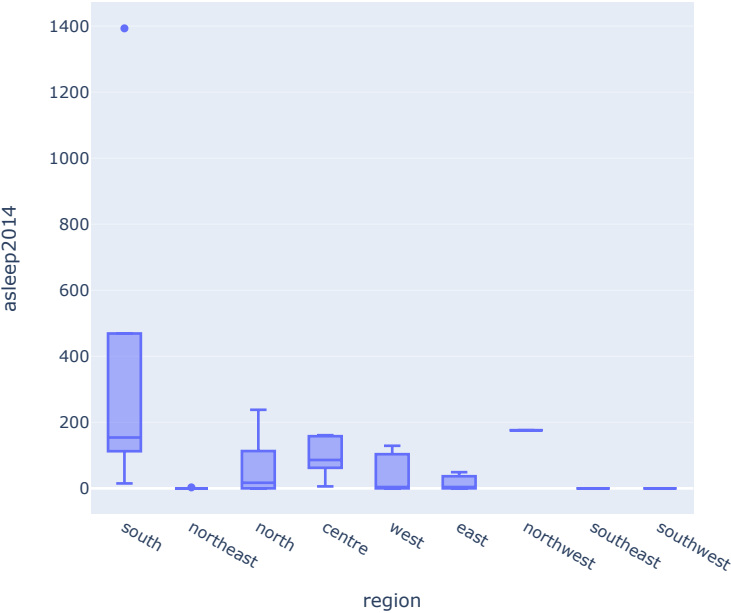
```
In [33]:  
fig = px.box(df, x='region', y='wrongside2014')  
fig.show()
```



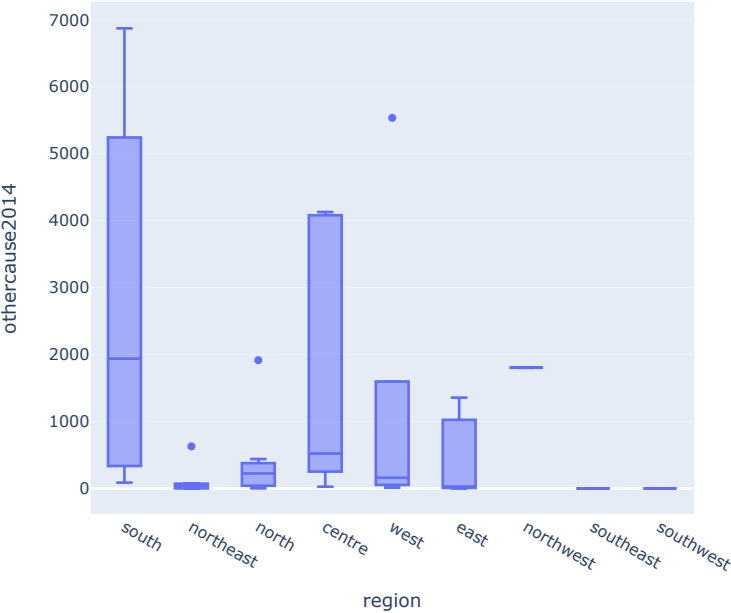
```
In [34]:  
fig = px.box(df, x='region', y='signalavoid2014')  
fig.show()
```



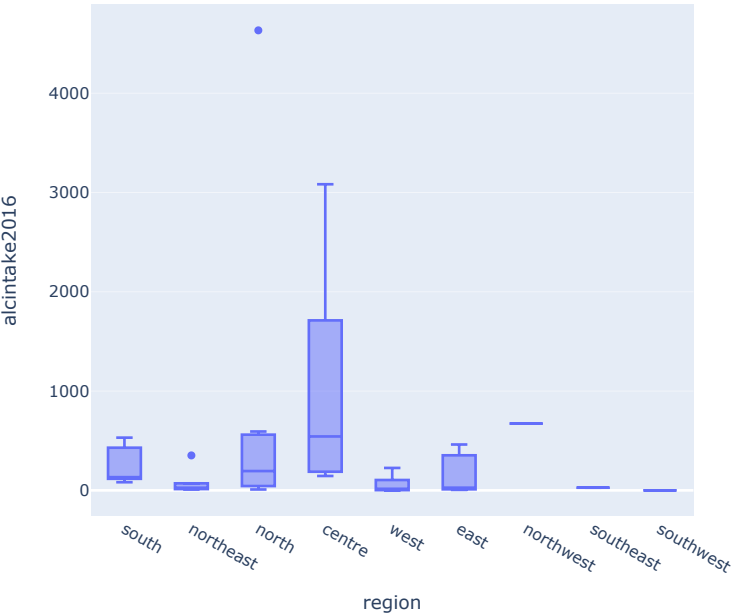

```
In [35]:  
fig = px.box(df, x='region', y='asleep2014')  
fig.show()
```



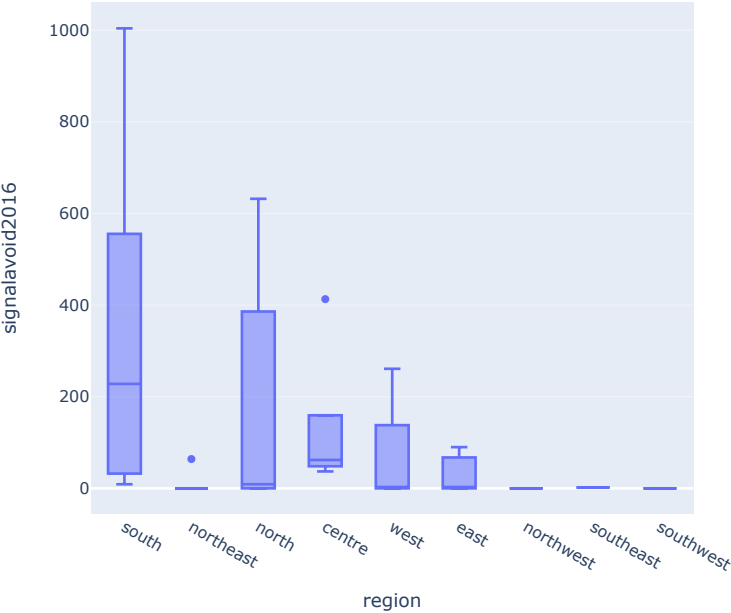
```
In [36]:  
fig = px.box(df, x='region', y='othercause2014')  
fig.show()
```



```
In [37]:  
fig = px.box(df, x='region', y='alcintake2016')  
fig.show()
```

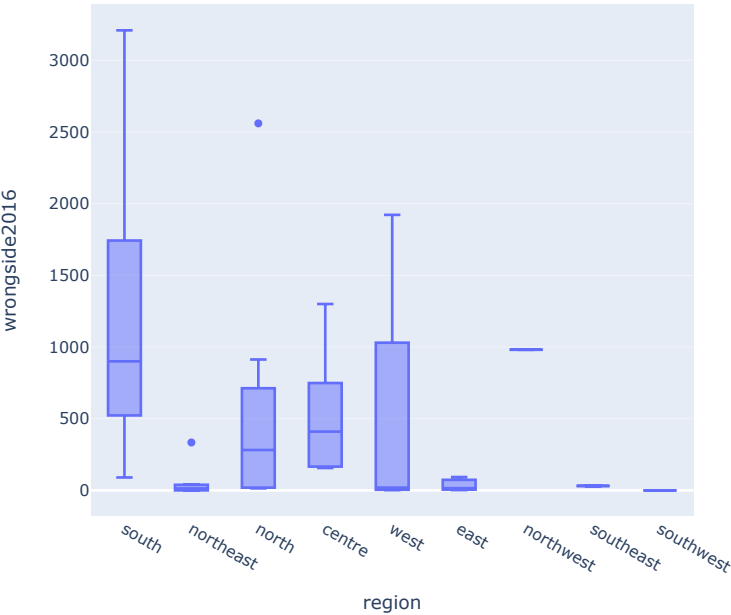


```
In [38]:  
fig = px.box(df, x='region', y='signalavoid2016')  
fig.show()
```

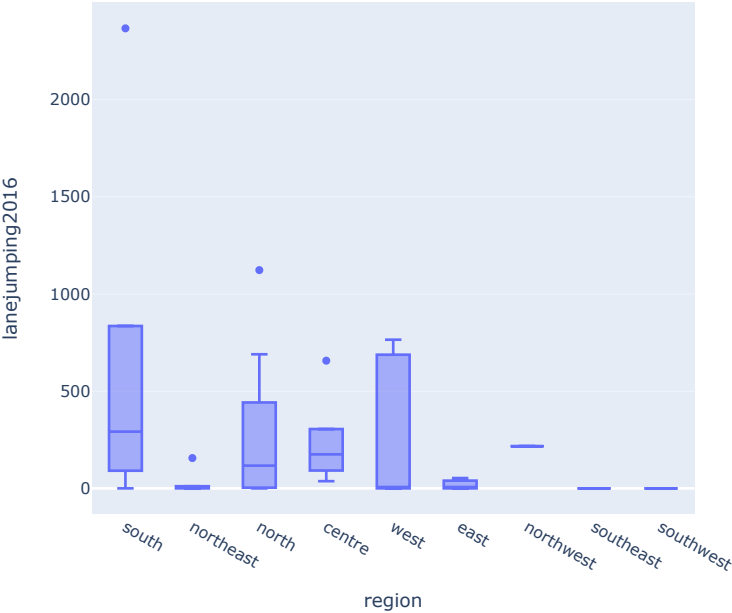


In [39]:

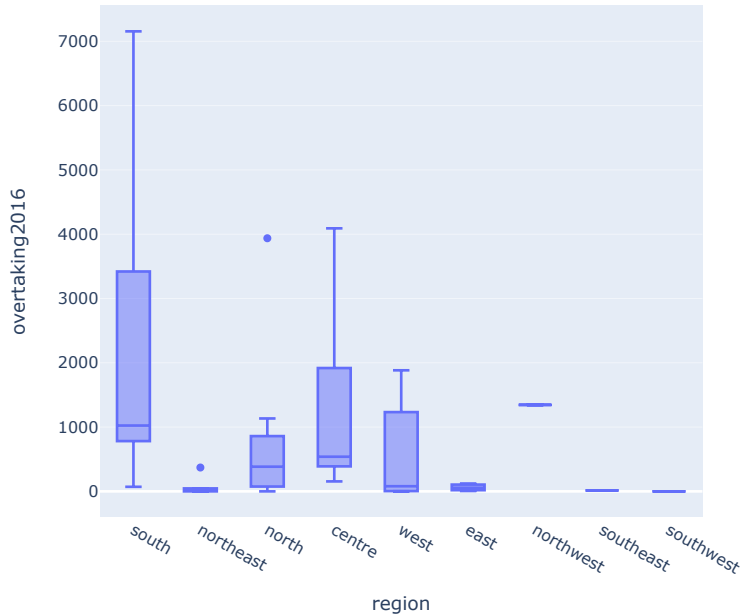
```
fig = px.box(df, x='region', y='wrongside2016')
fig.show()
```



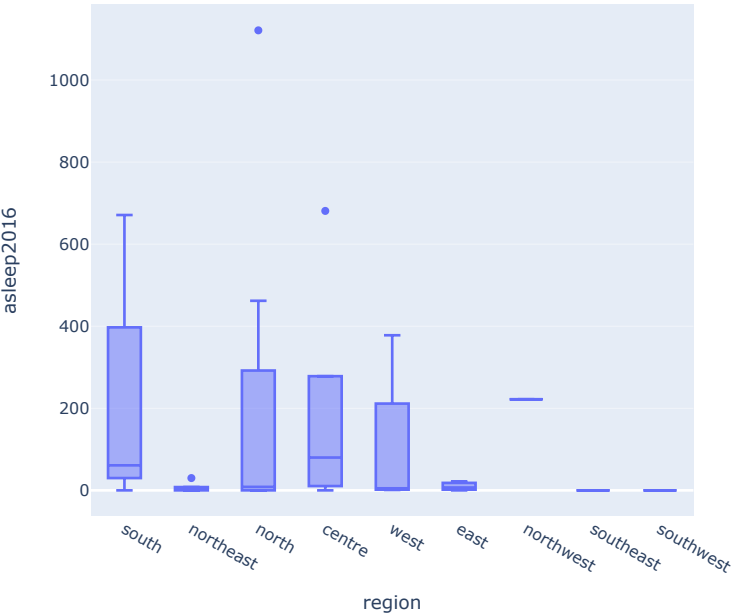
```
In [40]:  
fig = px.box(df, x='region', y='lanejumping2016')  
fig.show()
```



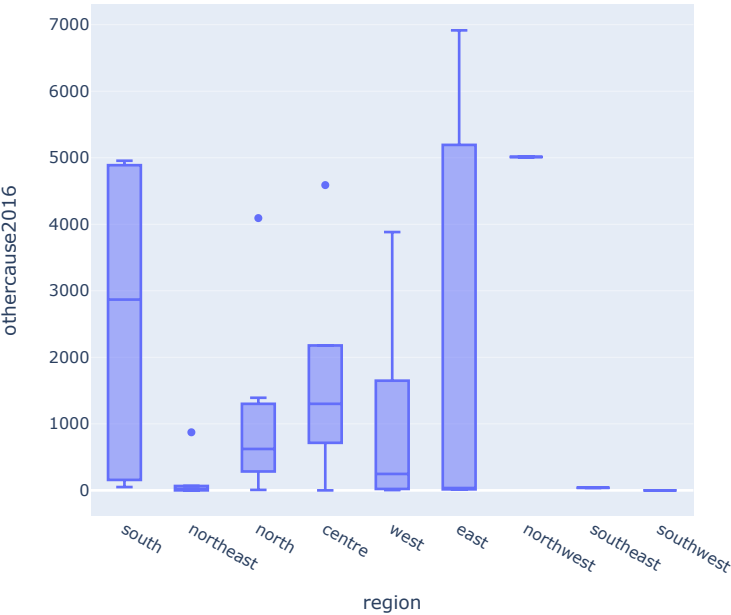
```
In [41]:  
fig = px.box(df, x='region', y='overtaking2016')  
fig.show()
```



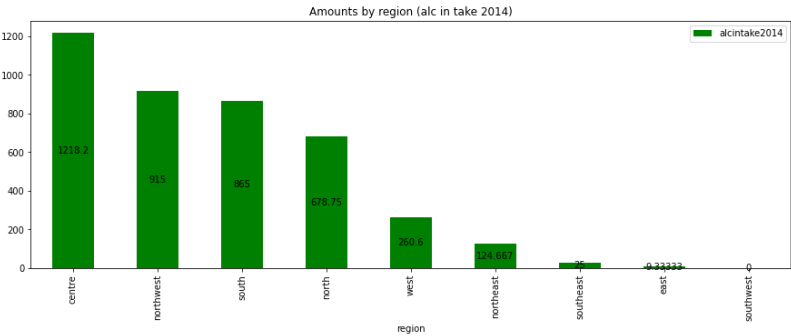
```
In [42]:  
fig = px.box(df, x='region', y='asleep2016')  
fig.show()
```




```
In [43]:  
fig = px.box(df, x='region', y='othercause2016')  
fig.show()
```

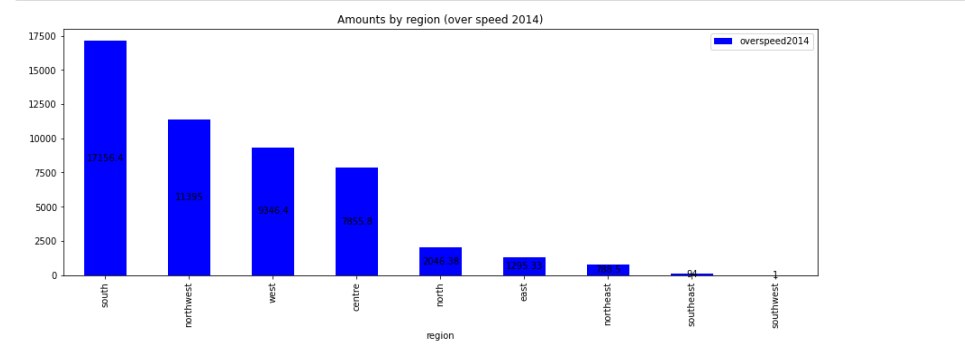


```
In [44]:  
q=pd.pivot_table(df,index='region',values='alcintake2014')  
q=q.sort_values(by='alcintake2014',ascending=False)  
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (alc in take 2014)', color="green")  
p.bar_label(p.containers[0], label_type='center')  
plt.show()
```



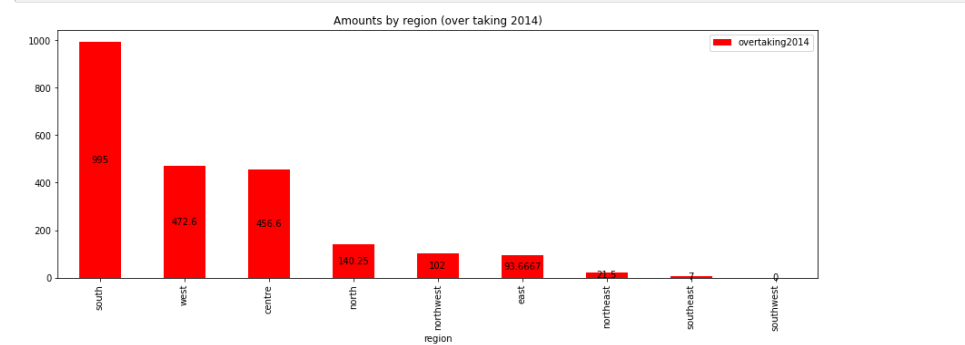
In [45]:

```
q=pd.pivot_table(df,index='region',values='overspeed2014')
q=q.sort_values(by='overspeed2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (over speed 2014)', color="blue")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



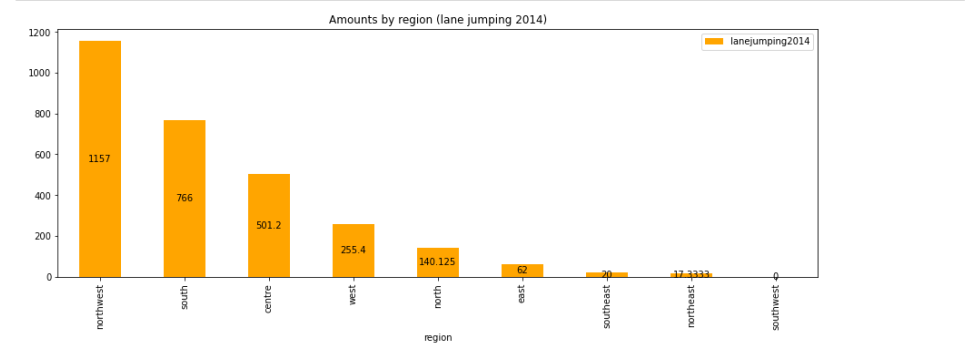
In [46]:

```
q=pd.pivot_table(df,index='region',values='overtaking2014')
q=q.sort_values(by='overtaking2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (over taking 2014)', color="red")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



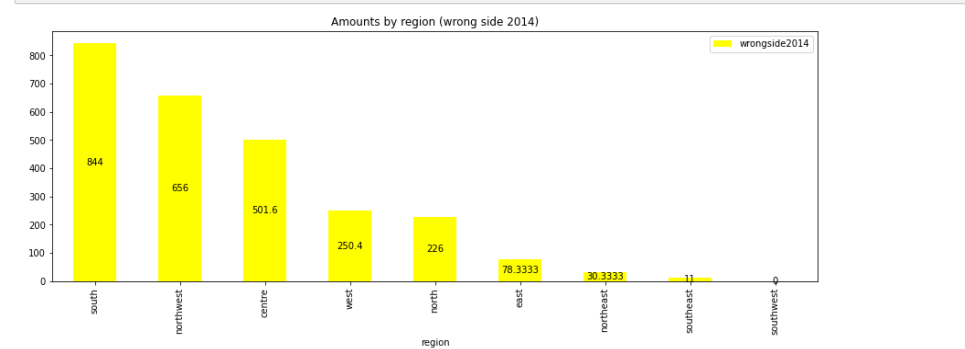
In [47]:

```
q=pd.pivot_table(df,index='region',values='lanejumping2014')
q=q.sort_values(by='lanejumping2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (lane jumping 2014)', color="orange")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



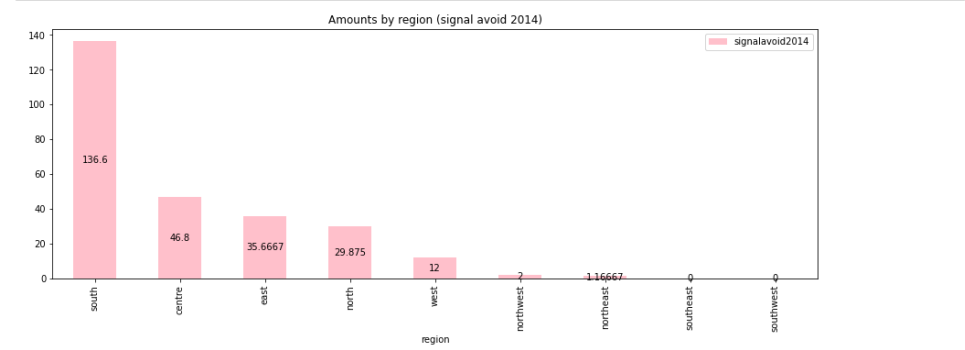
In [48]:

```
q=pd.pivot_table(df,index='region',values='wrongside2014')
q=q.sort_values(by='wrongside2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (wrong side 2014)', color="yellow")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



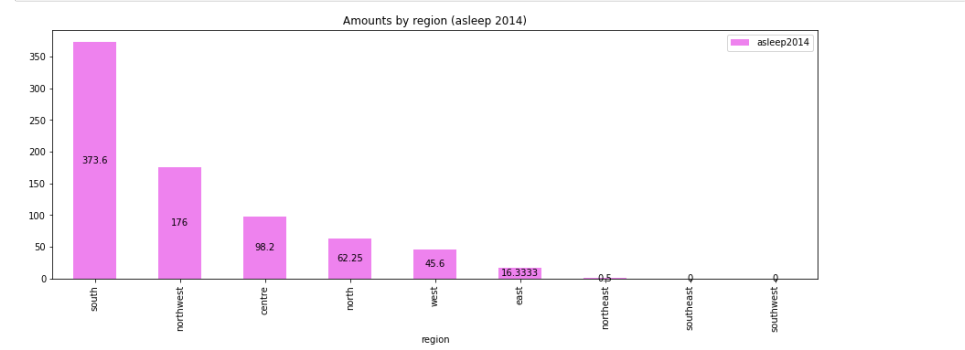
In [49]:

```
q=pd.pivot_table(df,index='region',values='signalavoid2014')
q=q.sort_values(by='signalavoid2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (signal avoid 2014)', color="pink")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```

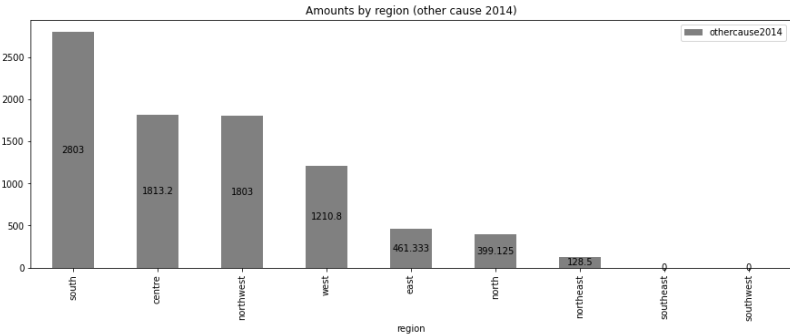


In [50]:

```
q=pd.pivot_table(df,index='region',values='asleep2014')
q=q.sort_values(by='asleep2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (asleep 2014)', color="violet")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```

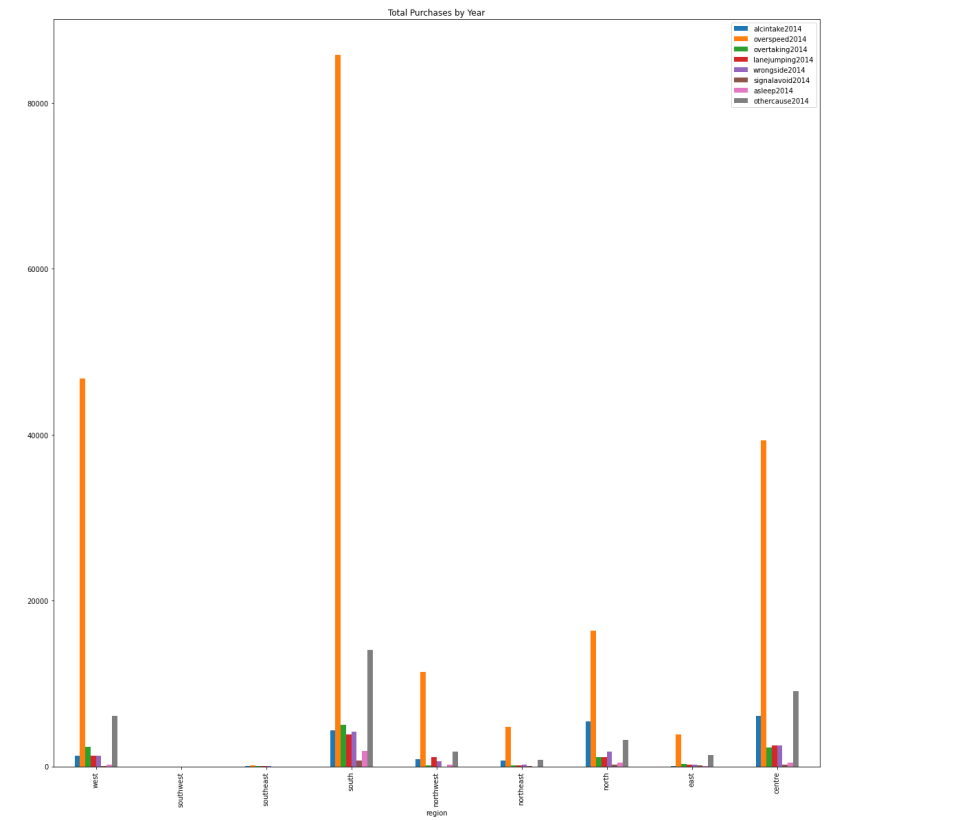


```
In [51]:
q=pd.pivot_table(df,index='region',values='othercause2014')
q=q.sort_values(by='othercause2014',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (other cause 2014)', color="grey")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```

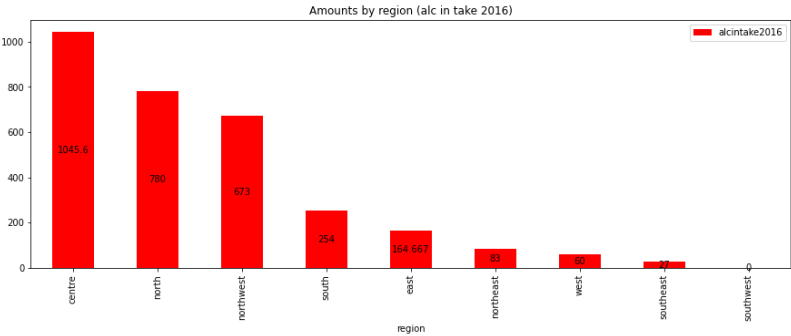


In [52]:

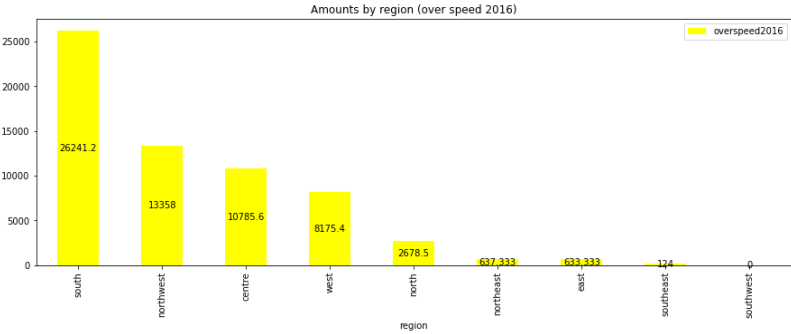
```
plt.rcParams['figure.figsize'] = (20, 20)
fig= df.groupby('region')[['alcintake2014','overspeed2014','overtaking2014', 'lanejumping2014']]
fig=fig.sort_values(by='region',ascending=False)
fig.plot(kind='bar',title='Total Purchases by Year')
plt.show()
```



```
In [53]:
q=pd.pivot_table(df,index='region',values='alcintake2016')
q=q.sort_values(by='alcintake2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (alc in take 2016)', color="red")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```

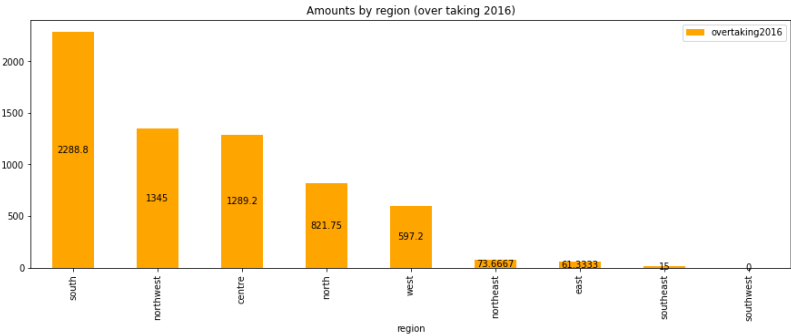


```
In [54]:
q=pd.pivot_table(df,index='region',values='overspeed2016')
q=q.sort_values(by='overspeed2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (over speed 2016)', color="yellow")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



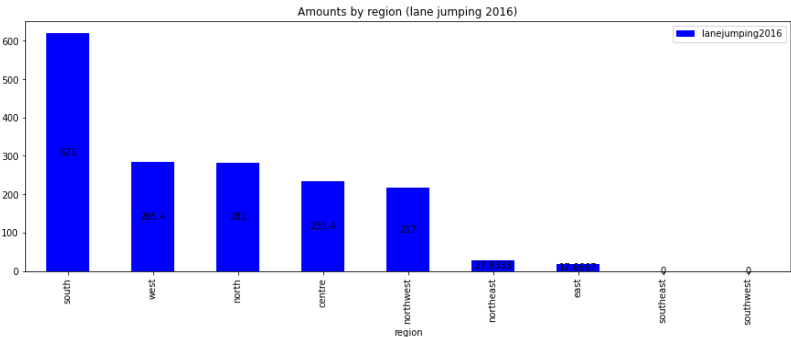
In [55]:

```
q=pd.pivot_table(df,index='region',values='overtaking2016')
q=q.sort_values(by='overtaking2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (over taking 2016)', color="orange")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



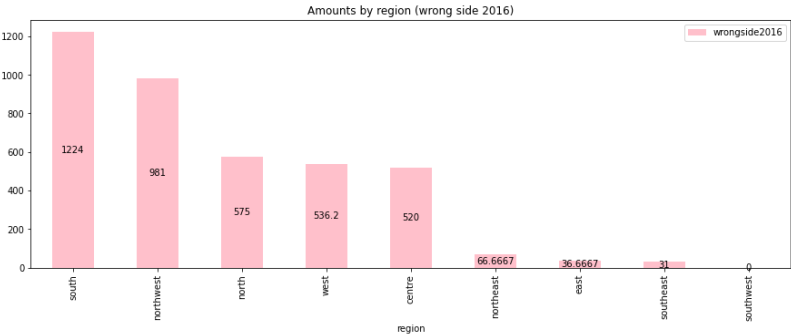
In [56]:

```
q=pd.pivot_table(df,index='region',values='lanejumping2016')
q=q.sort_values(by='lanejumping2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (lane jumping 2016)', color="blue")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



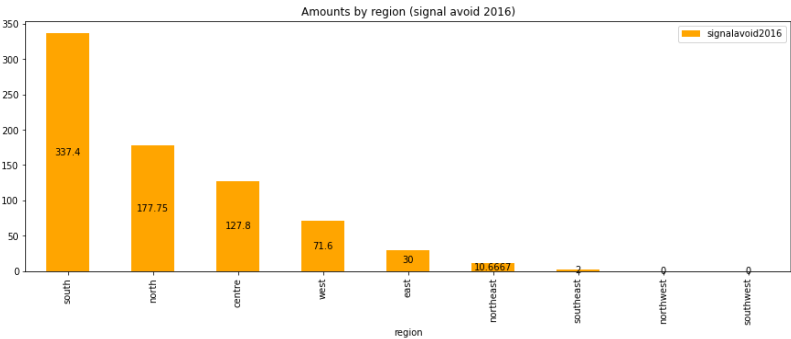
In [57]:

```
q=pd.pivot_table(df,index='region',values='wrongside2016')
q=q.sort_values(by='wrongside2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (wrong side 2016)', color="pink")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



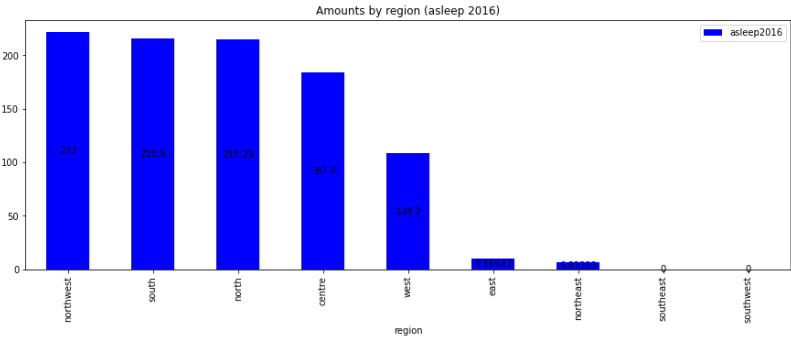
In [58]:

```
q=pd.pivot_table(df,index='region',values='signalavoid2016')
q=q.sort_values(by='signalavoid2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (signal avoid 2016)', color="orange")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



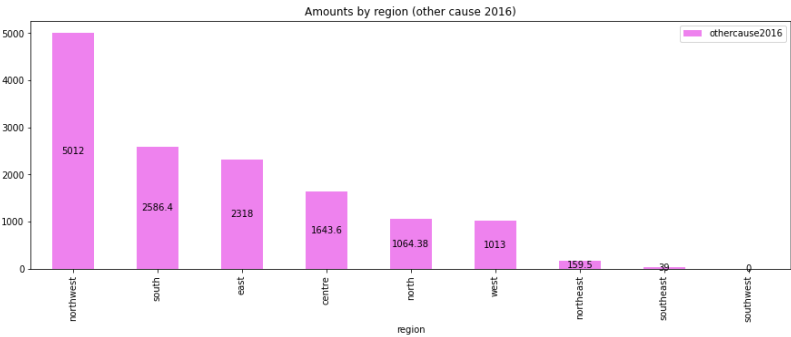
In [59]:

```
q=pd.pivot_table(df,index='region',values='asleep2016')
q=q.sort_values(by='asleep2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (asleep 2016)', color="blue")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



In [60]:

```
q=pd.pivot_table(df,index='region',values='othercause2016')
q=q.sort_values(by='othercause2016',ascending=False)
p=q.plot(kind='bar',figsize=(15,5),title='Amounts by region (other cause 2016)', color="violet")
p.bar_label(p.containers[0], label_type='center')
plt.show()
```



```
In [61]:
plt.rcParams['figure.figsize'] = (20, 20)
fig2= df.groupby('region')[['alcintake2016','overspeed2016','overtaking2016', 'lanejumping20
fig2=fig2.sort_values(by='region',ascending=False)
fig2.plot(kind='bar',title='Total Purchases by Year')
plt.show()
```

