P R O J E C T

Click on below link to download dataset:
Dataset link (https://info.stackoverflowsolutions.com/rs/719-EMH-566/images/stack-overflow-developer-survey-2022.zip)

importing libraries

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pycountry
import plotly.express as px
from wordcloud import WordCloud
import warnings; warnings.filterwarnings('ignore')
```

reading dataset

In [2]:
```
survey_df = pd.read_csv('sods2022/survey_results_public.csv')      # answers
```

In [3]:
```
schema_df = pd.read_csv('sods2022/survey_results_schema.csv')      # questions
```

In [4]:
```
1 survey_df.head()
```

Out[4]:

	ResponseId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel	LearnCode	LearnCodeOnline	LearnCodeCoursesCert	YearsC
0	1	None of these	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1	2	I am a developer by profession	Employed, full-time	Fully remote	Hobby;Contribute to open-source projects	NaN	NaN	NaN	NaN	N
2	3	I am not primarily a developer, but I write co...	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Books / Physical media;Friend or family member...	Technical documentation;Blogs;Programming Game...	NaN	
3	4	I am a developer by profession	Employed, full-time	Fully remote	I don't code outside of work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;School (i.e., Universit...	NaN	NaN	
4	5	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Other online resources (e.g., videos, blogs, f...	Technical documentation;Blogs;Stack Overflow;O...	NaN	

5 rows × 79 columns

In [5]:
```
1 schema_df.head()
```

Out[5]:

	qid	qname	question	force_resp	type	selector
0	QID16	S0	Hel...	False	DB	TB
1	QID12	MetaInfo	Browser Meta Info	False	Meta	Browser
2	QID1	S1	<span style="font-size:22px; font-family: aria...	False	DB	TB
3	QID2	MainBranch	Which of the following options best describes ...	True	MC	SAVR
4	QID296	Employment	Which of the following best describes your cur...	False	MC	MAVR

preprocessing

we need column *qname* as index of `schema_df` DataFrame

In [6]:
```
schema_df.set_index('qname', inplace=True)
```

In [7]:
```
schema_df.head(2)
```

Out[7]:

qname	qid	question	force_resp	type	selector
S0	QID16	Hel...	False	DB	TB
MetaInfo	QID12	Browser Meta Info	False	Meta	Browser

In [ ]:
```

```

In [8]:
```
schema_df = schema_df.question
```

After deletion

In [10]:
```
1 schema_df
```

Out[10]:
```
qname
S0            <div><span style="font-size:19px;"><strong>Hel...
MetaInfo                                      Browser Meta Info
S1            <span style="font-size:22px; font-family: aria...
MainBranch    Which of the following options best describes ...
Employment    Which of the following best describes your cur...
                                    ...
Frequency_2   Interacting with people outside of your immedi...
Frequency_3   Encountering knowledge silos (where one indivi...
TrueFalse_1   Are you involved in supporting new hires durin...
TrueFalse_2   Do you use learning resources provided by your...
TrueFalse_3   Does your employer give you time to learn new ...
Name: question, Length: 79, dtype: object
```

In [11]:
```
1 schema_df.index
```

Out[11]:
```
Index(['S0', 'MetaInfo', 'S1', 'MainBranch', 'Employment', 'RemoteWork',
       'CodingActivities', 'S2', 'EdLevel', 'LearnCode', 'LearnCodeOnline',
       'LearnCodeCoursesCert', 'YearsCode', 'YearsCodePro', 'DevType',
       'OrgSize', 'PurchaseInfluence', 'BuyNewTool', 'Country', 'Currency',
       'CompTotal', 'CompFreq', 'S3', 'Language', 'Database', 'Platform',
       'Webframe', 'MiscTech', 'ToolsTech', 'NEWCollabTools', 'OpSys',
       'VersionControlSystem', 'VCInteraction', 'VCHosting',
       'OfficeStackAsync', 'OfficeStackSync', 'Blockchain', 'S4', 'NEWSOSites',
       'SOVisitFreq', 'SOAccount', 'SOPartFreq', 'SOComm', 'S5', 'Age',
       'Gender', 'Trans', 'Sexuality', 'Ethnicity', 'Accessibility',
       'MentalHealth', 'S6', 'TBranch', 'ICorPM', 'WorkExp', 'Knowledge',
       'Frequency', 'TimeSearching', 'TimeAnswering', 'Onboarding',
       'ProfessionalTech', 'SOTeamsUsage', 'TrueFalse', 'S7', 'SurveyLength',
       'SurveyEase', 'Knowledge_1', 'Knowledge_2', 'Knowledge_3',
       'Knowledge_4', 'Knowledge_5', 'Knowledge_6', 'Knowledge_7',
       'Frequency_1', 'Frequency_2', 'Frequency_3', 'TrueFalse_1',
       'TrueFalse_2', 'TrueFalse_3'],
      dtype='object', name='qname')
```

In [12]:

```
1 survey_df
```

Out[12]:

	ResponseId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel	LearnCode	LearnCodeOnline	LearnCodeCourse
0	1	None of these	NaN	NaN	NaN	NaN	NaN	NaN	
1	2	I am a developer by profession	Employed, full-time	Fully remote	Hobby;Contribute to open-source projects	NaN	NaN	NaN	
2	3	I am not primarily a developer, but I write co...	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Books / Physical media;Friend or family member...	Technical documentation;Blogs;Programming Game...	
3	4	I am a developer by profession	Employed, full-time	Fully remote	I don't code outside of work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;School (i.e., Universit...	NaN	
4	5	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Other online resources (e.g., videos, blogs, f...	Technical documentation;Blogs;Stack Overflow;O...	
...	...	...	...	...	...	...	...	...	
73263	73264	I am a developer by profession	Employed, full-time	Fully remote	Freelance/ contract work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;Other online resources ...	Technical documentation;Blogs;Written Tutorial...	U
73264	73265	I am a developer by profession	Employed, full-time	Full in-person	Hobby	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Other online resources (e.g., videos, blogs, f...	Technical documentation;Blogs;Written Tutorial...	Coursera;Udemy;U
73265	73266	I am not primarily a developer, but I write co...	Employed, full-time	Hybrid (some remote, some in-person)	Hobby;School or academic work	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;Other online resources ...	Technical documentation;Programming Games;Stac...	Udemy;Codecademy;Pluralsigl

	ResponseId	MainBranch	Employment	RemoteWork	CodingActivities	EdLevel	LearnCode	LearnCodeOnline	LearnCodeCourse
73266	73267	I am a developer by profession	Employed, full-time	Hybrid (some remote, some in-person)	Hobby	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;On the job training	NaN	
73267	73268	I used to be a developer by profession, but no...	Independent contractor, freelancer, or self-em...	Fully remote	Hobby;Contribute to open-source projects;Boots...	Bachelor's degree (B.A., B.S., B.Eng., etc.)	Books / Physical media;Friend or family member...	Technical documentation;Blogs;Programming Game...	Udemy;Plura

73268 rows × 79 columns

plot function

In [13]:

```
def custom_plot(series, plot_height=15, plot_width=5,
                y_label_font_size=13.5,
                title = '', title_font_size=15,
                percent_font_size=14,
                color = 'light:#59C1BD'):

    # create figure to display plot
    plt.figure( figsize=(plot_width, plot_height) )

    # to hide square of the plot
    custom_params = {
                        "axes.spines.bottom": False,
                        "axes.spines.right": False,
                        "axes.spines.left"  : False,
                        "axes.spines.top": False
                    }

    sns.set_theme(style="white", rc=custom_params)

    # creating different shades of colors(color palette) of size series length
    # pal stores rgb values for different color shades
    pal = sns.color_palette(color, len(series)) # light:#5A9

    # argsort return indices of elements according to sorting order..
    # means lowest number will be indexed as 0, and so on
    # rank stores rank of series whr highest count value comes first
    # using this rank to assign color shades to diffrnt bars in plot
    rank = series.argsort().argsort()

    ax = sns.barplot(x = series.values, y=series.index,
                     #palette='PuBuGn_r'
                     #order=series.sort_values('Growth').State,
                     palette=np.array(pal[::])[rank]
                    )

    # to calculate percentage
    s = series.values.sum()

    for rect in ax.patches:
        x_value = rect.get_width()
        y_value = (rect.get_y() + rect.get_height() / 2)
        space = 0

        # calculating percentage and assigning to variable label
        label = "{:.2f}%".format( (100*x_value/s))

```

```
        # to display percentage value on bar
        plt.annotate(
            text=label,                     # Use `label` as label
            xy=((x_value/2)-5, y_value),    # Place label at end of the bar, xy argument..maybe coordinate
            xytext=(space, 0),              # Horizontally shift label by `space`
            textcoords="offset points",     # Interpret `xytext` as offset in points
            va='center',                    # Vertically center label
            color = 'white',
            #ha='center',
            weight='bold', size=percent_font_size
        )

    plt.title('\n'+title+'\n',
              fontdict=
              {
                  "color": 'black',
                  "weight":'bold',
                  "size":title_font_size
              }
             )


    plt.yticks(size=y_label_font_size)#, weight='bold')
    plt.xticks([], []) # to hide xticks

    f_dict={"color": 'black',"weight":'bold',  "size":15}
    plt.figtext(.74, .042, "Total Responses: {}".format(s),
                fontdict = f_dict);
```

what is your main branch..?

In [14]:
```
schema_df.MainBranch
```

Out[14]:
```
'Which of the following options best describes you today? Here, by "developer" we mean "someone who writes code." <b>*</b>'
```

In [15]:
```
title = 'Which of the following options best describes you today?'
```

In [16]:
```
mb = survey_df.MainBranch.value_counts()
```

In [17]:
```
mb
```

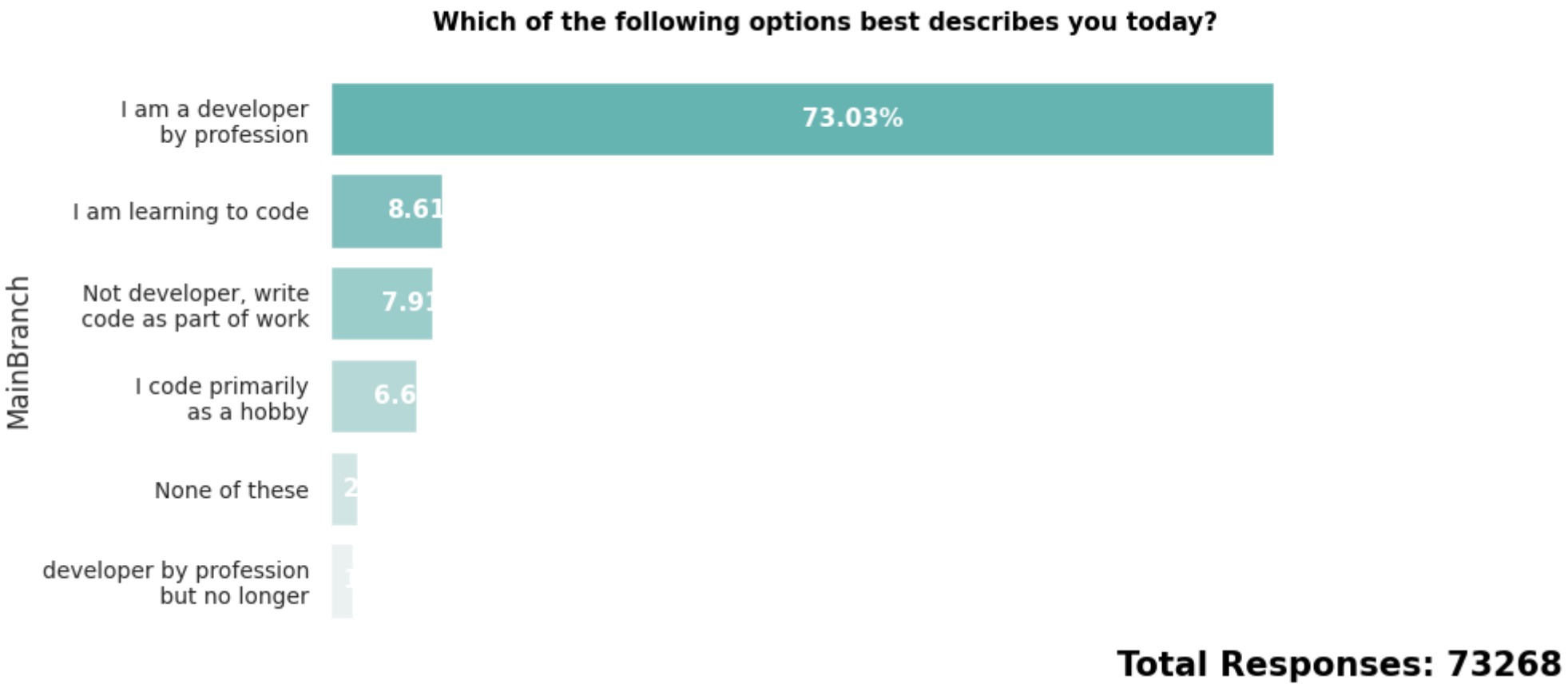
Out[17]:
```
MainBranch
I am a developer by profession                                          53507
I am learning to code                                                    6309
I am not primarily a developer, but I write code sometimes as part of my work     5794
I code primarily as a hobby                                              4865
None of these                                                            1497
I used to be a developer by profession, but no longer am                 1296
Name: count, dtype: int64
```

In [18]:
```
def MainBranch_ylabel_text_process(s):
    if s == 'I am not primarily a developer, but I write code sometimes as part of my work':
        return 'Not developer, write\n code as part of work'
    elif s == 'I used to be a developer by profession, but no longer am':
        return 'developer by profession\n but no longer'
    elif s == 'I am a developer by profession':
        return 'I am a developer\n by profession'
    elif s == 'I code primarily as a hobby':
        return 'I code primarily\n as a hobby'
    else:
        return s
```

In [19]:
```
survey_df.MainBranch = survey_df.MainBranch.apply(MainBranch_ylabel_text_process, )
```

In [20]:
```
mb = survey_df.MainBranch.value_counts()
```

In [21]:
```
custom_plot(mb, plot_height=4.5, plot_width=8, y_label_font_size=10, title=title, title_font_size=11,
            percent_font_size=11)
```



How old is the average professional developer..?

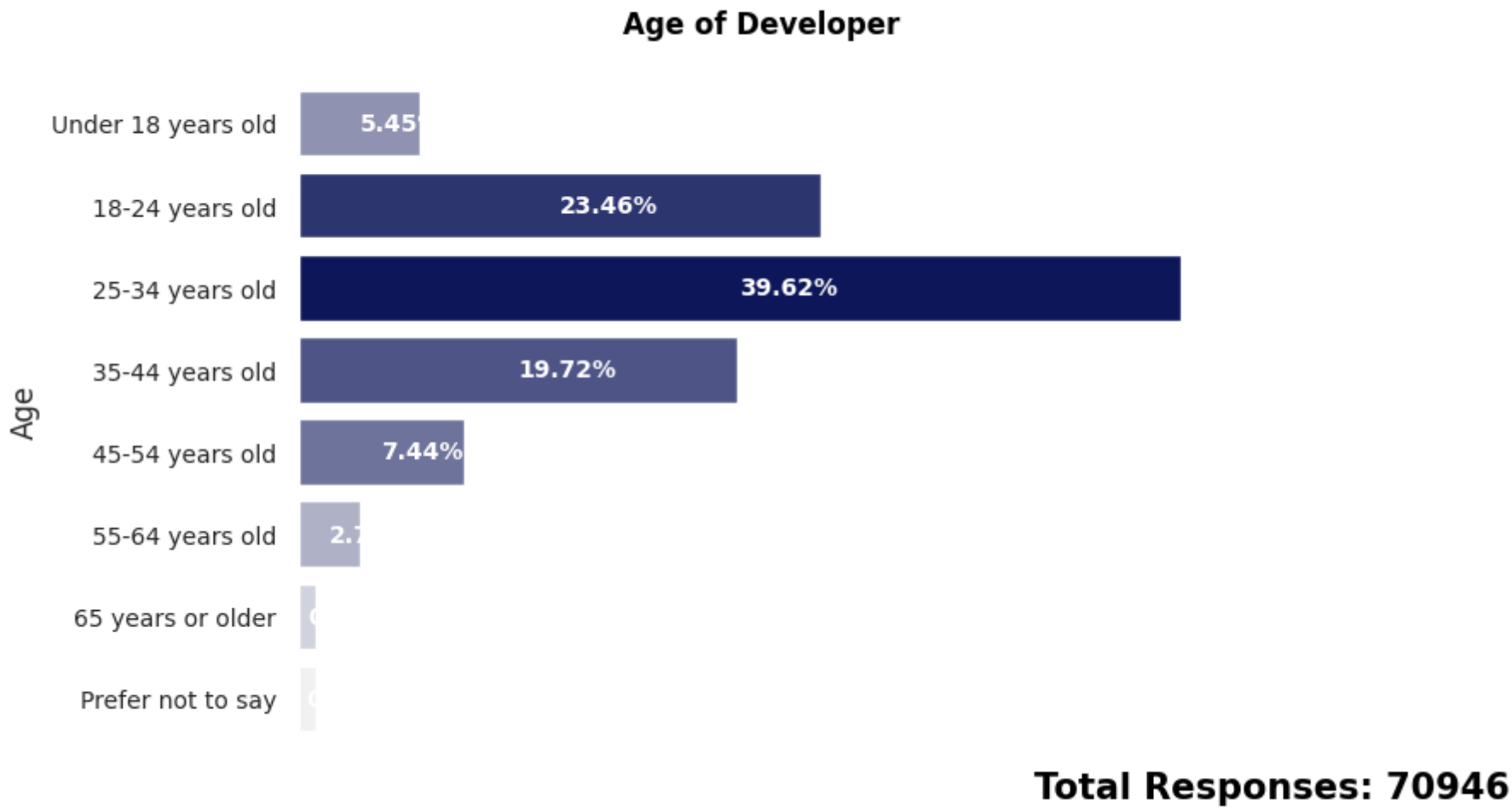
In [22]:

```
reorder_list = ['Under 18 years old', '18-24 years old',
                '25-34 years old', '35-44 years old',
                 '45-54 years old', '55-64 years old',
                 '65 years or older', 'Prefer not to say']

age = survey_df.Age.value_counts().reindex(reorder_list)
age
```

Out[22]:

```
Age
Under 18 years old     3866
18-24 years old       16646
25-34 years old       28112
35-44 years old       13988
45-54 years old        5281
55-64 years old        1978
65 years or older       554
Prefer not to say       521
Name: count, dtype: int64
```

In [25]:
```
custom_plot( age, plot_height=5, color='light:#000C66', title='Age of Developer',
            percent_font_size=10, y_label_font_size=10, title_font_size=12, plot_width=7)
```



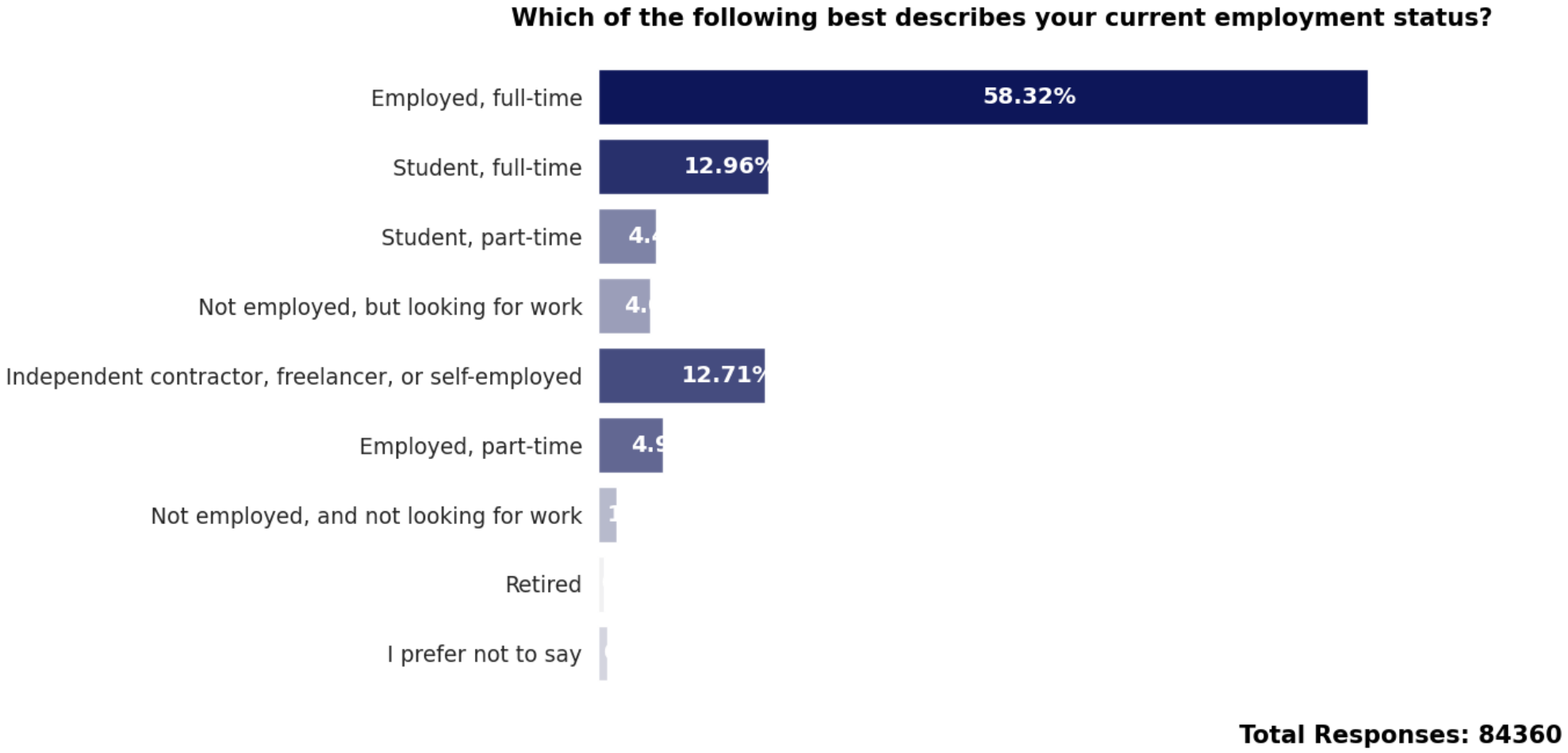
Employment status of an employee

In [28]:

```
def colum_expand( s ):
    d = {}

    for t in s.dropna().values:
        for i in t.split(';'):
            if i in d.keys():
                d[i] += 1
            else:
                d[i] = 1

    return pd.Series(d)
```

In [32]:

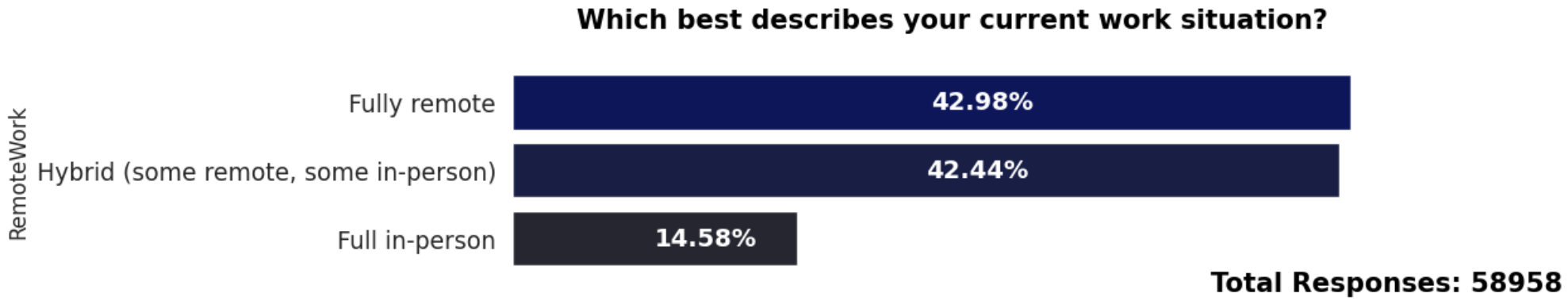
```
emp = colum_expand(survey_df.Employment)

custom_plot(emp, plot_height=7, color='light:#000C66',
            title=schema_df.Employment, plot_width=9)
```



mode of working of employee(remote/hybrid)

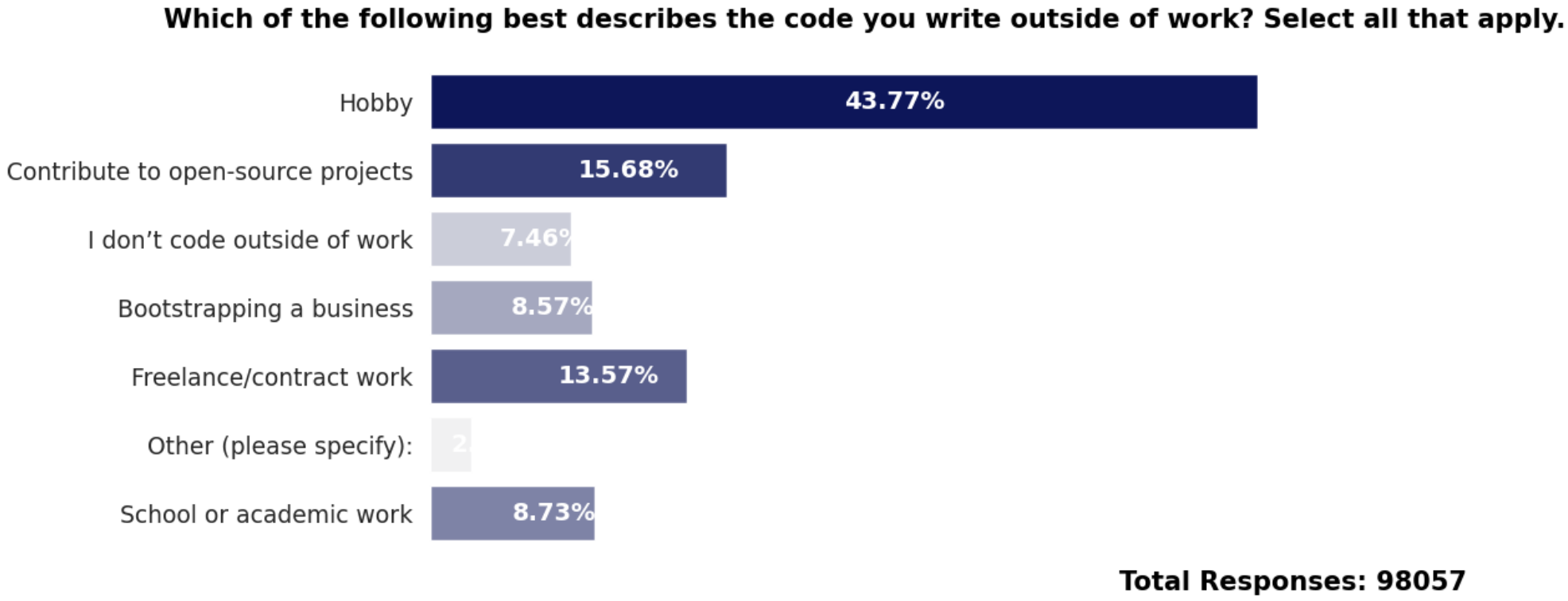
In [36]:

```
remote_work = survey_df.RemoteWork.value_counts()

custom_plot(remote_work, plot_height=2.1, plot_width=9
            , color='dark:#000C66', title=schema_df.RemoteWork)
```



how many of you write code outside of your work

In [41]:
```
coding_act = colum_expand(survey_df.CodingActivities)

custom_plot(coding_act, plot_height=5, plot_width=9,
            color='light:#000C66', title=schema_df.CodingActivities)
```



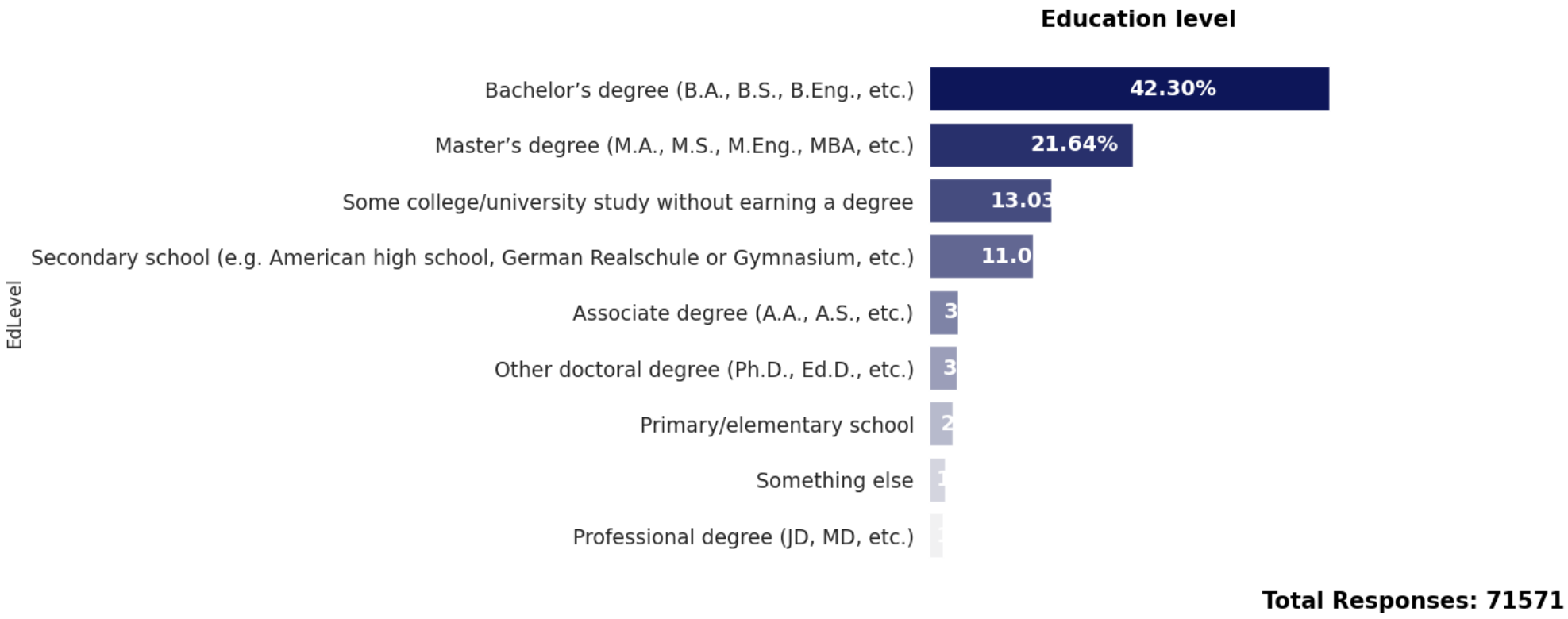
What is your highest level of formal education..?

In [42]:
```
# EdLevel
schema_df.EdLevel
```

Out[42]: 'Which of the following best describes the highest level of formal education that you’ve completed? *'

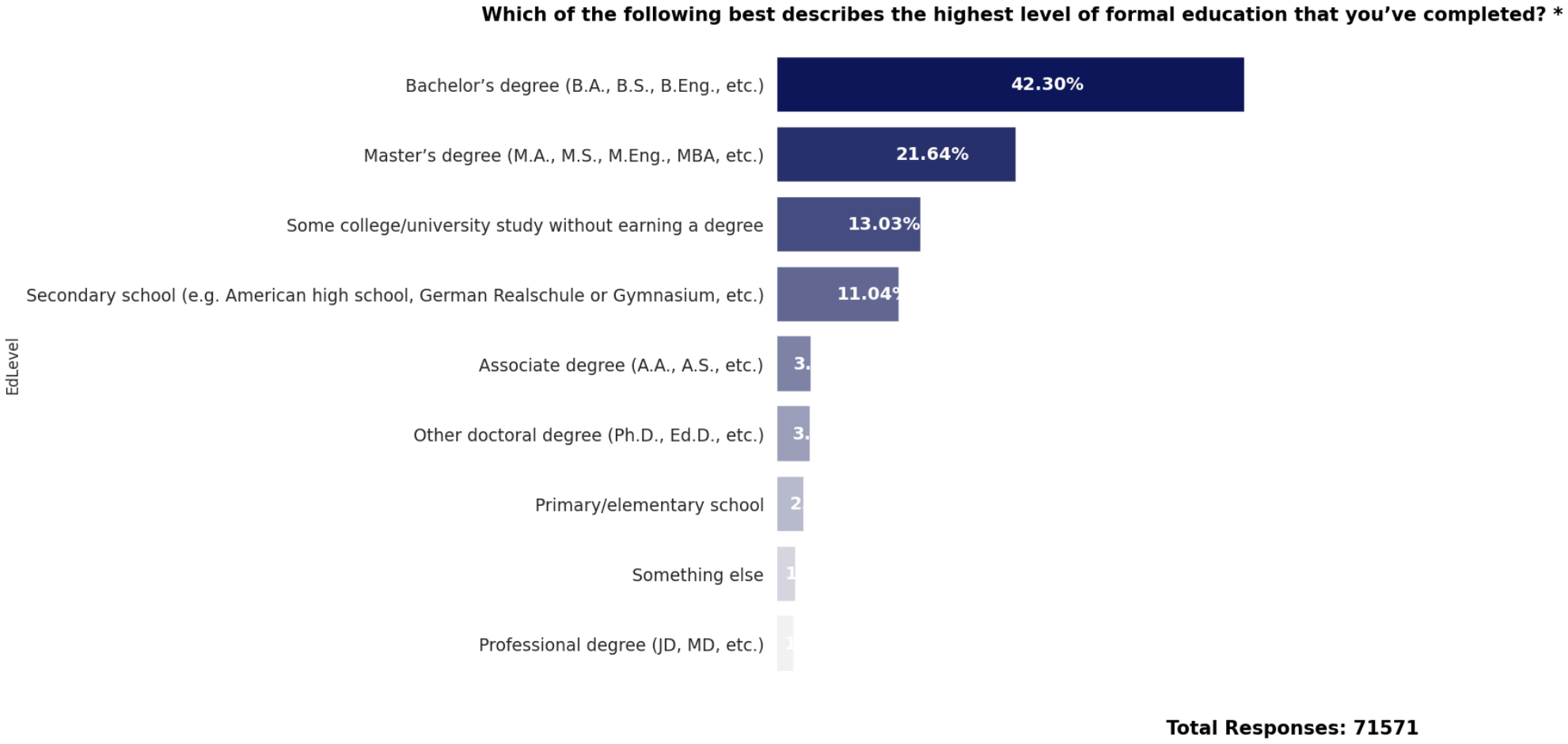
In [43]:
```
ed_level = survey_df.EdLevel.value_counts()
```

In [44]:
```
1 custom_plot(ed_level, plot_height=6, title='Education level', color='light:#000C66')
```



In [45]:

```
edu = survey_df.EdLevel.value_counts()

custom_plot(edu, plot_height=9, plot_width=7, color='light:#000C66',
            title=schema_df.EdLevel)
```



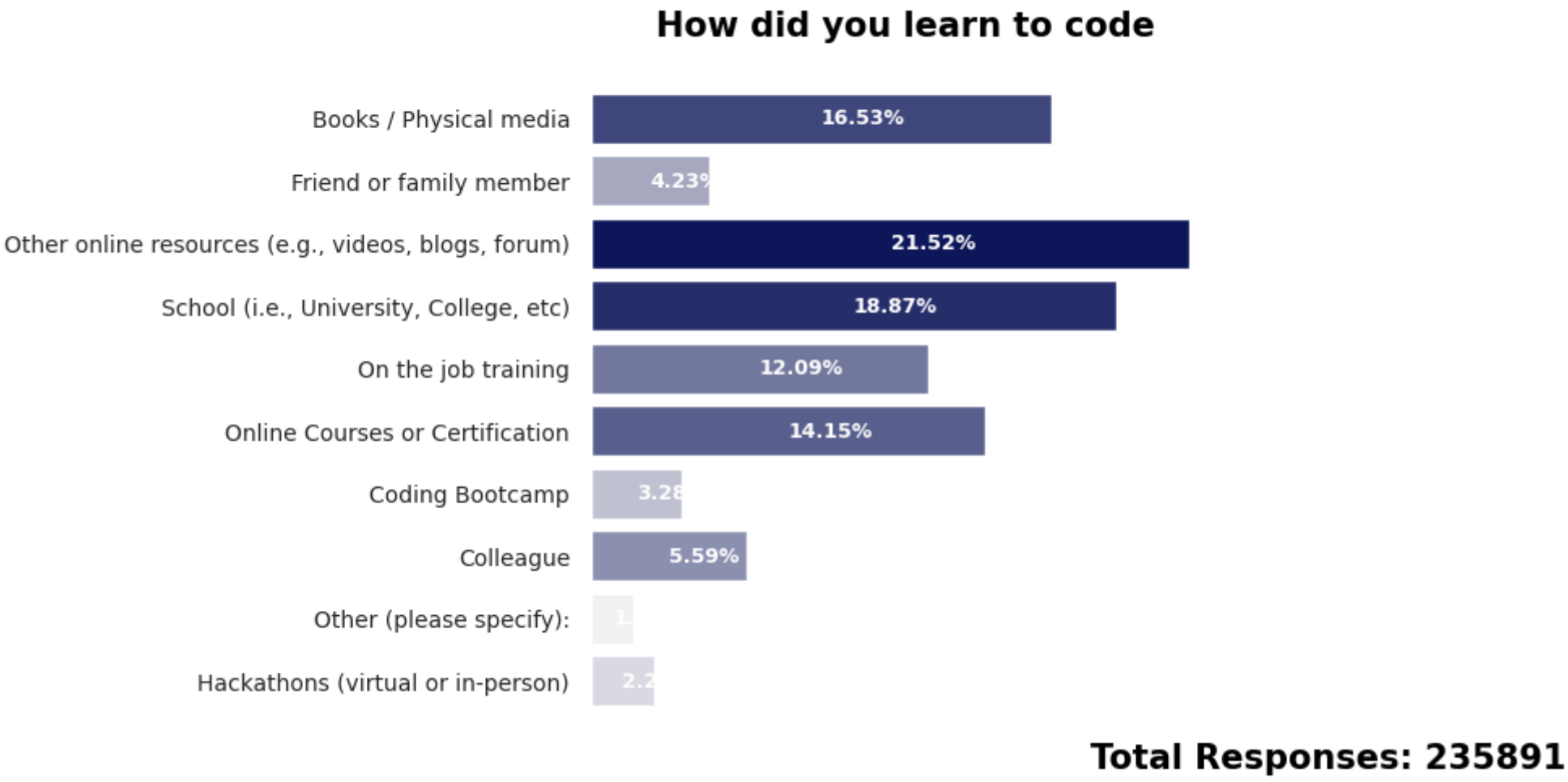
How did you learn to code

In [46]:
```
# LearnCode
schema_df.LearnCode
```

Out[46]: 'How did you learn to code? Select all that apply.'

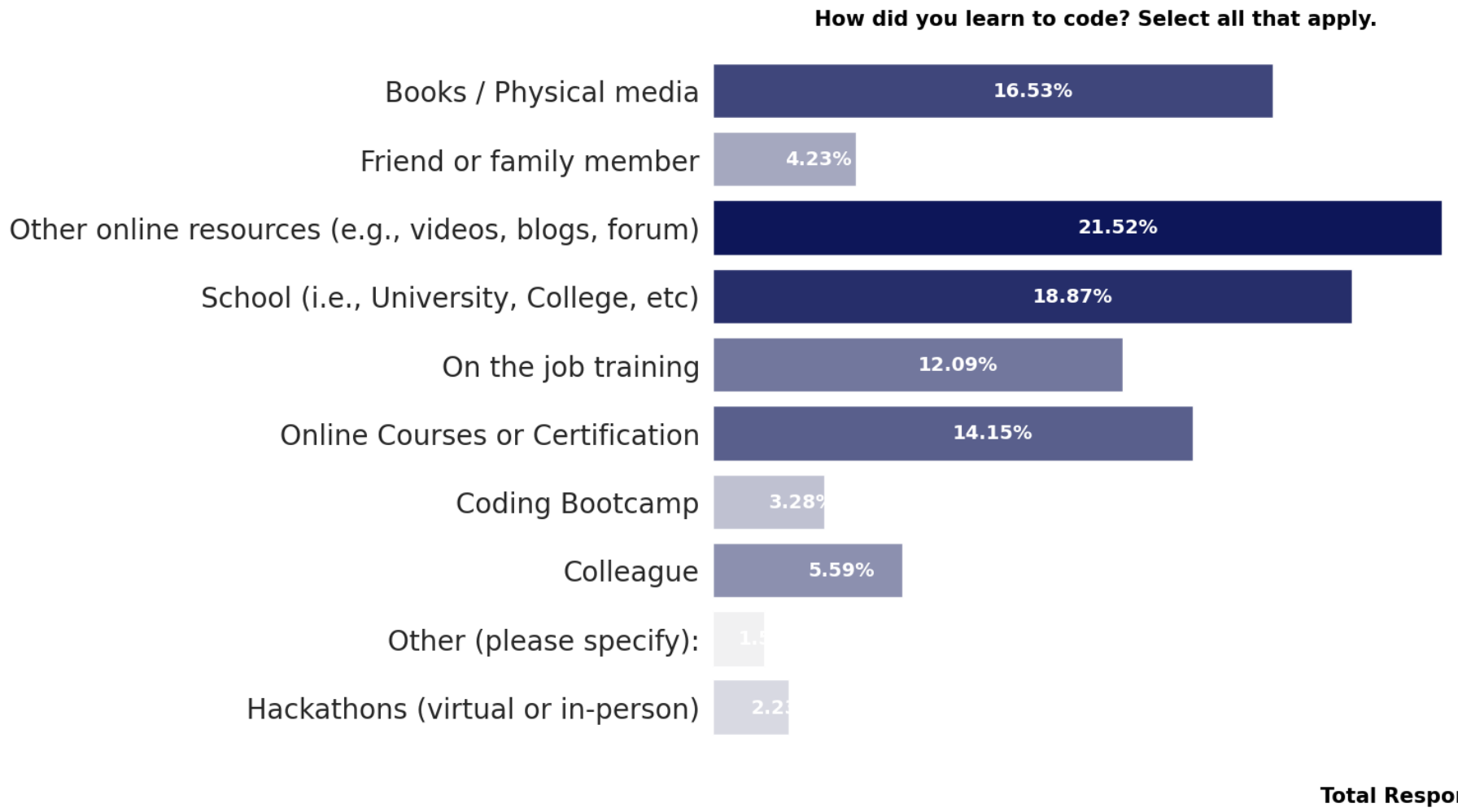
In [47]:
```
lc = colum_expand(survey_df.LearnCode)
```

In [48]:
```
custom_plot(lc, plot_height=5, color='light:#000C66', y_label_font_size=10,
            title='How did you learn to code', percent_font_size=9)
```



In [49]:

```
learn_code_data = colum_expand(survey_df.LearnCode)

custom_plot(learn_code_data, plot_height=9, plot_width=10,
            color='light:#000C66', title=schema_df.LearnCode,
            y_label_font_size=20)

```



What online resources do you use to learn to code?

In [50]:
```
# LearnCodeOnline
```

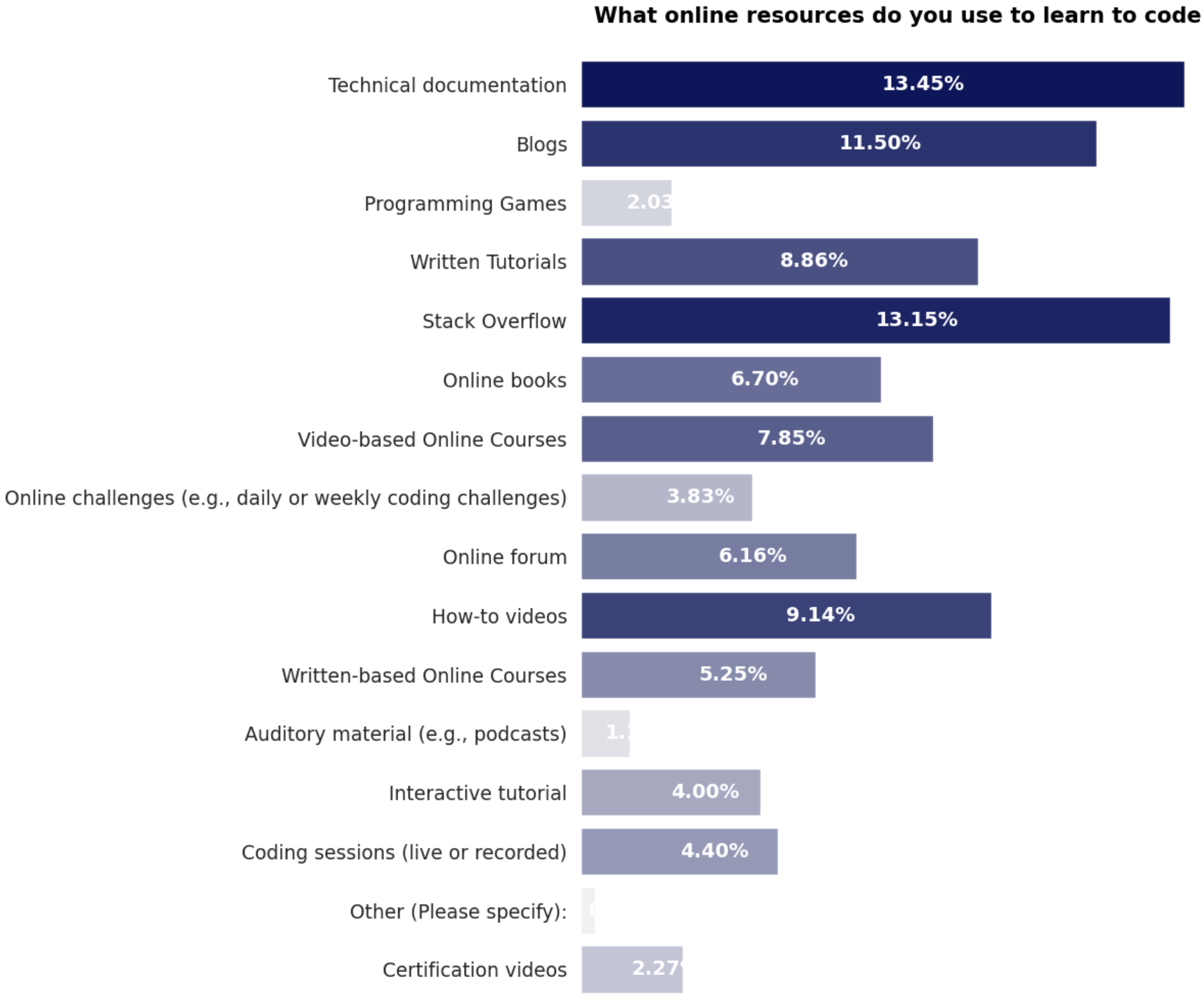
In [51]:
```
schema_df.LearnCodeOnline
```

Out[51]:
```
'What online resources do you use to learn to code? Select all that apply.'
```

In [52]:
```
lco = colum_expand(survey_df.LearnCodeOnline)
```

In [53]:

```
custom_plot(lco, plot_height=12, plot_width=8, color='light:#000C66',
            title='What online resources do you use to learn to code')
```

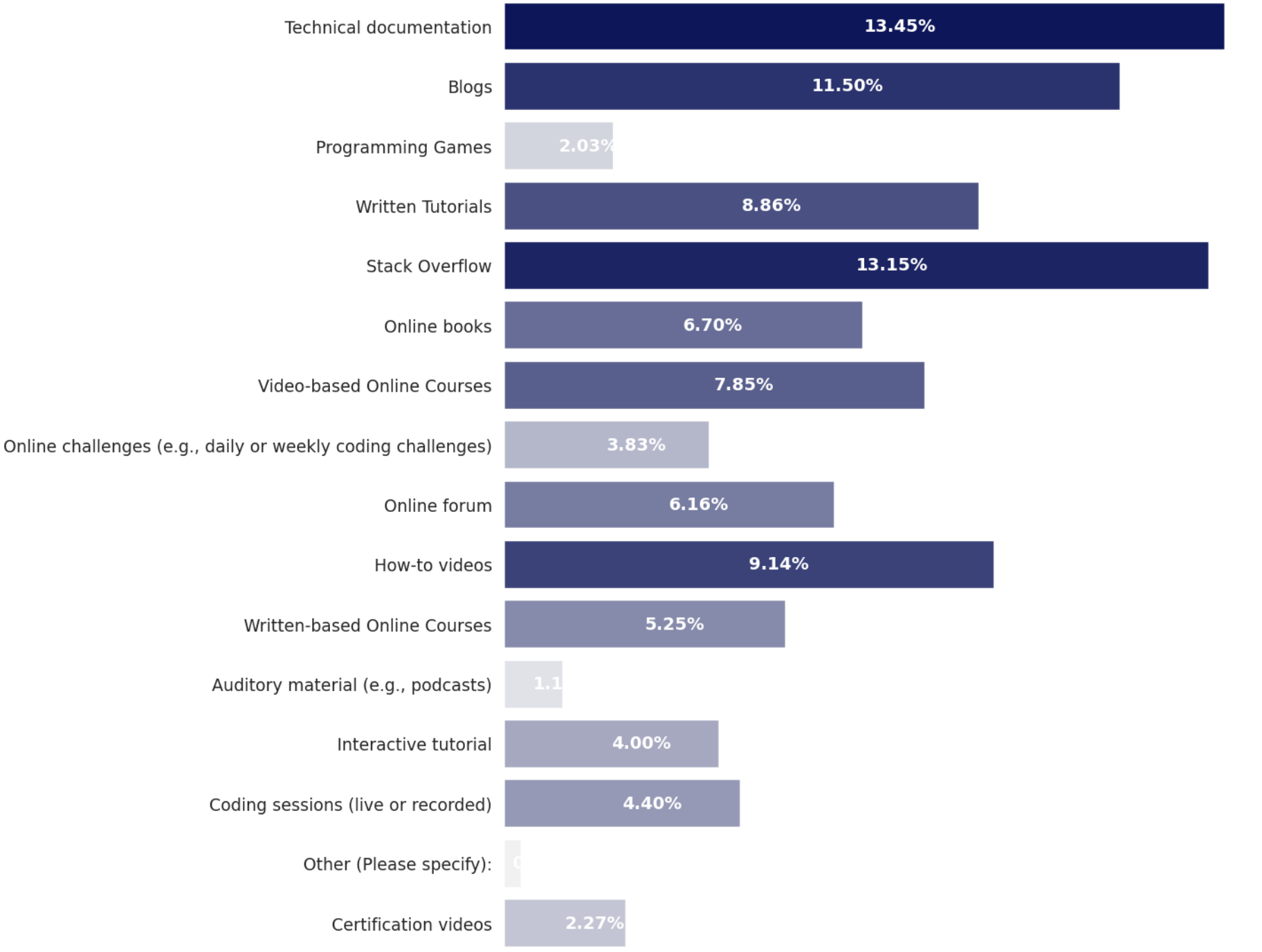


Total Responses: 332107

In [54]:

```
learn_code_online = colum_expand(survey_df.LearnCodeOnline)


custom_plot(learn_code_online, plot_height=14, plot_width=11,
            color='light:#000C66', title=schema_df.LearnCodeOnline)

```

What online resources do you use to learn to code? Select all that apply.



Total Responses: 332107

What online courses or certifications do you use to learn to code?

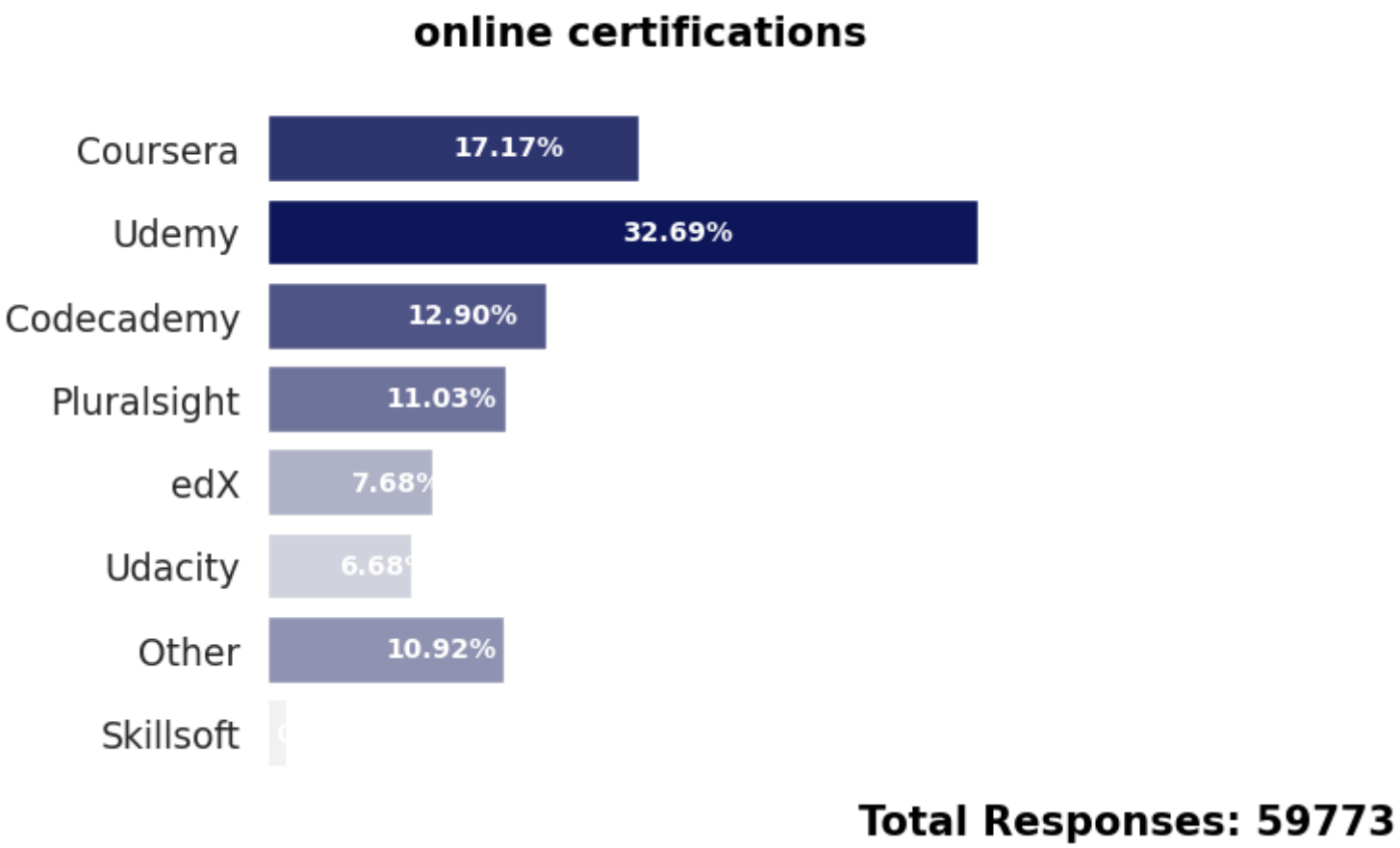
In [55]:
```
1 # LearnCodeCoursesCert
```

In [56]:
```
1 schema_df.LearnCodeCoursesCert
```

Out[56]:
```
'What online courses or certifications do you use to learn to code? Select all that apply.'
```

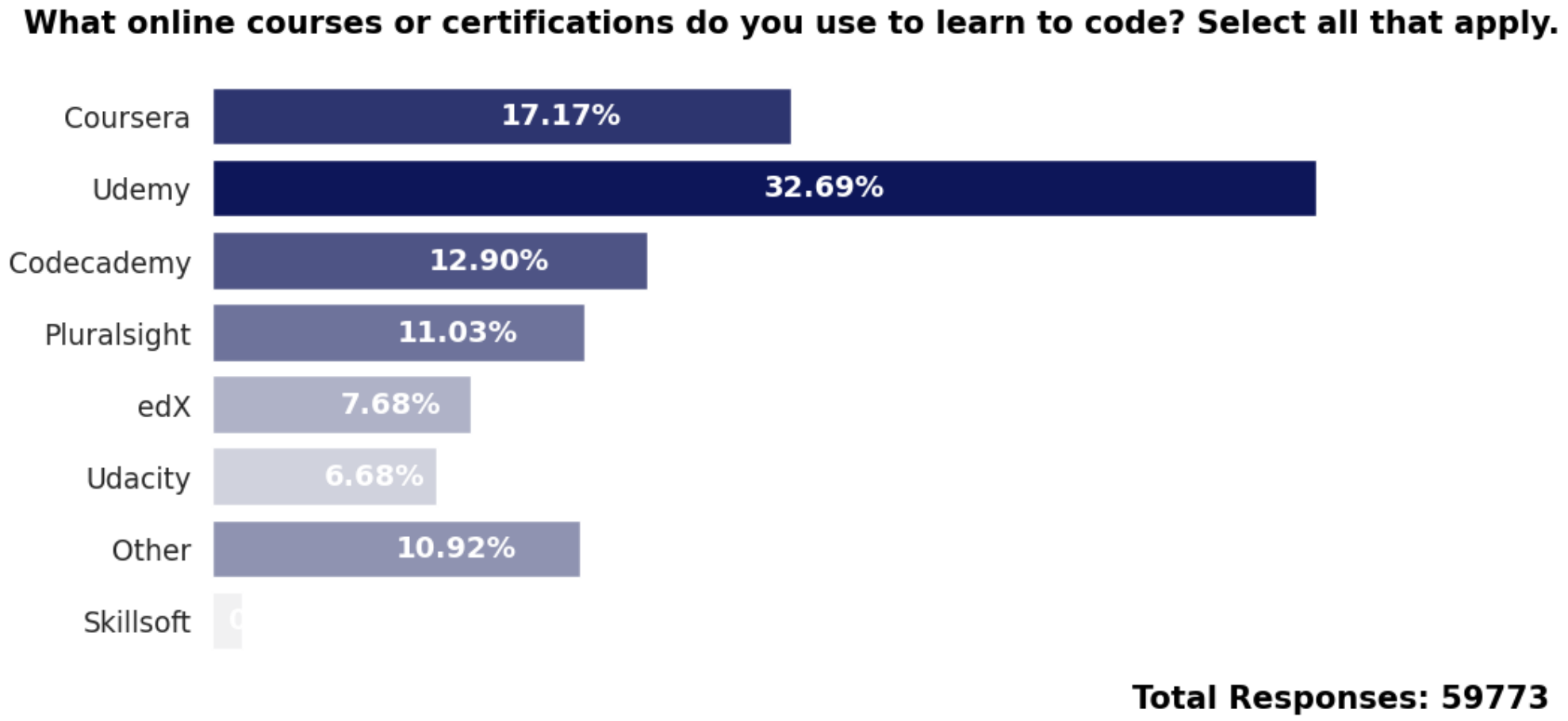
In [57]:
```
1 lccc = colum_expand(survey_df.LearnCodeCoursesCert)
```

In [58]:
```
custom_plot( lccc, plot_height=4.5, color='light:#000C66', title='online certifications',
            percent_font_size=10)
```



In [59]:

```
learn_code_cert = colum_expand(survey_df.LearnCodeCoursesCert)

custom_plot(learn_code_cert, plot_height=5, plot_width=10,
            color='light:#000C66', title=schema_df.LearnCodeCoursesCert)
```



how many years have you been coding in total (Including education)

In [63]:

```
def make_groups(s):
    try:
        s = int(s)
        if s > 0 and s < 5:
            return '0 to 5 years'
        if s > 5 and s < 10:
            return '5 to 10 years'
        if s > 10 and s < 15:
            return '10 to 15 years'
        if s > 15 and s < 20:
            return '15 to 20 years'
        if s > 20 and s < 25:
            return '20 to 25 years'
        if s > 25 and s < 30:
            return '25 to 30 years'
        if s > 30 and s < 35:
            return '30 to 35 years'
        if s > 35 and s < 40:
            return '35 to 40 years'
        if s > 40 and s < 45:
            return '40 to 45 years'
        if s > 45 and s < 50:
            return '45 to 50 years'
    except (TypeError, ValueError):
        pass

```

In [61]:

```
schema_df.YearsCode
```

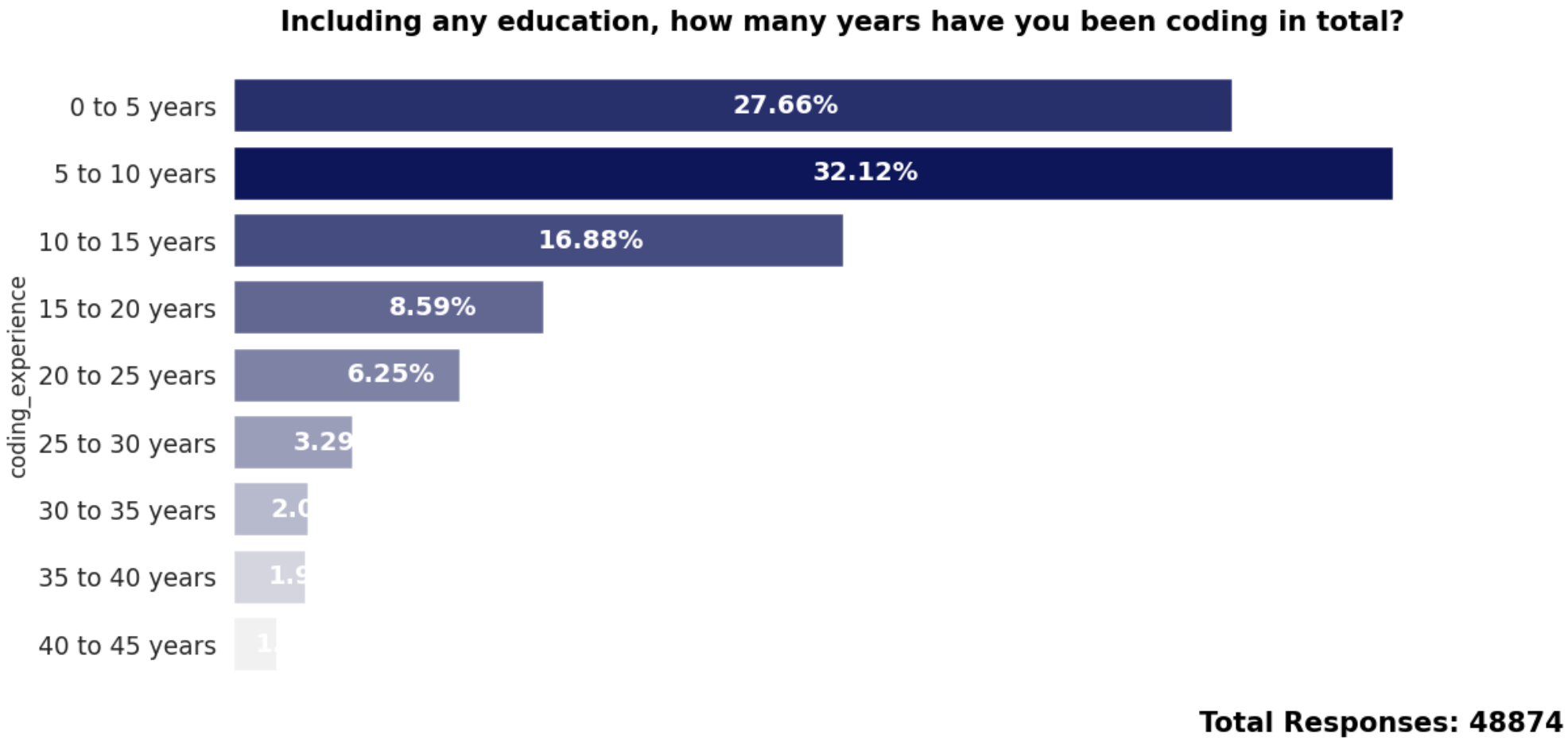
Out[61]: 'Including any education, how many years have you been coding in total?'

In [64]:

```
survey_df['coding_experience'] = survey_df.YearsCode.apply(make_groups)

```

In [65]:

```
reorder_list = ['0 to 5 years', '5 to 10 years', '10 to 15 years',
                '15 to 20 years', '20 to 25 years', '25 to 30 years',
                '30 to 35 years','35 to 40 years', '40 to 45 years']

ce = survey_df.coding_experience.value_counts().reindex(reorder_list)

custom_plot(ce, plot_height=6, plot_width=12,
            title=schema_df.YearsCode, color='light:#000C66')
```



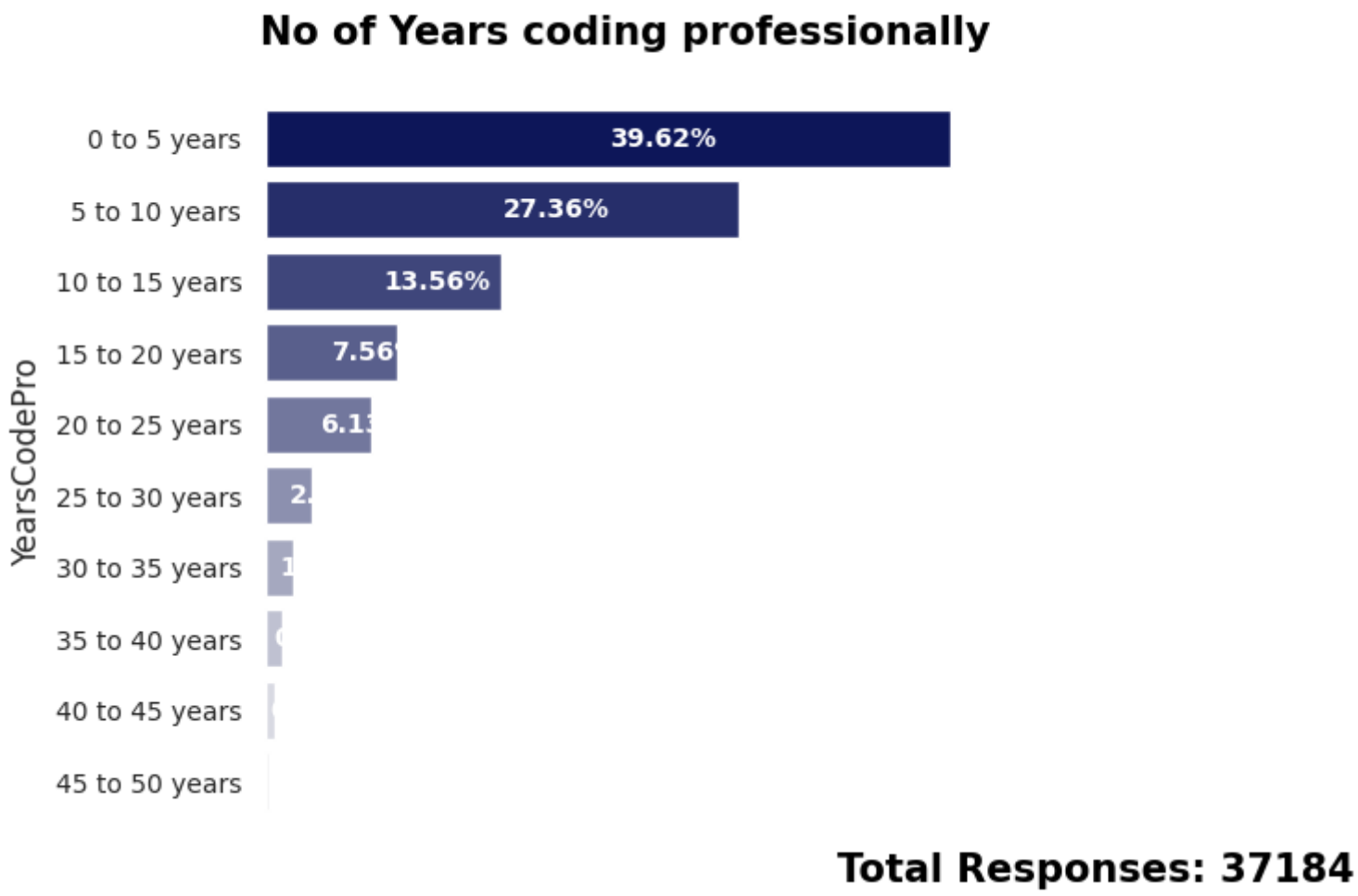
how many years have you been coding in total (not Including education)

In [67]:
```
1 ycp = survey_df.YearsCodePro.apply( make_groups ).value_counts()
```

In [68]:
```
1 ycp
```

Out[68]:
```
YearsCodePro
0 to 5 years      14734
5 to 10 years     10173
10 to 15 years     5042
15 to 20 years     2811
20 to 25 years     2279
25 to 30 years      986
30 to 35 years      579
35 to 40 years      358
40 to 45 years      166
45 to 50 years       56
Name: count, dtype: int64
```

In [69]:
```
custom_plot( ycp, plot_height=5, color='light:#000C66', title='No of Years coding professionally',
            y_label_font_size=10, percent_font_size=10)
```



what kind of developer you are..?

In [72]:
```
schema_df.DevType
```

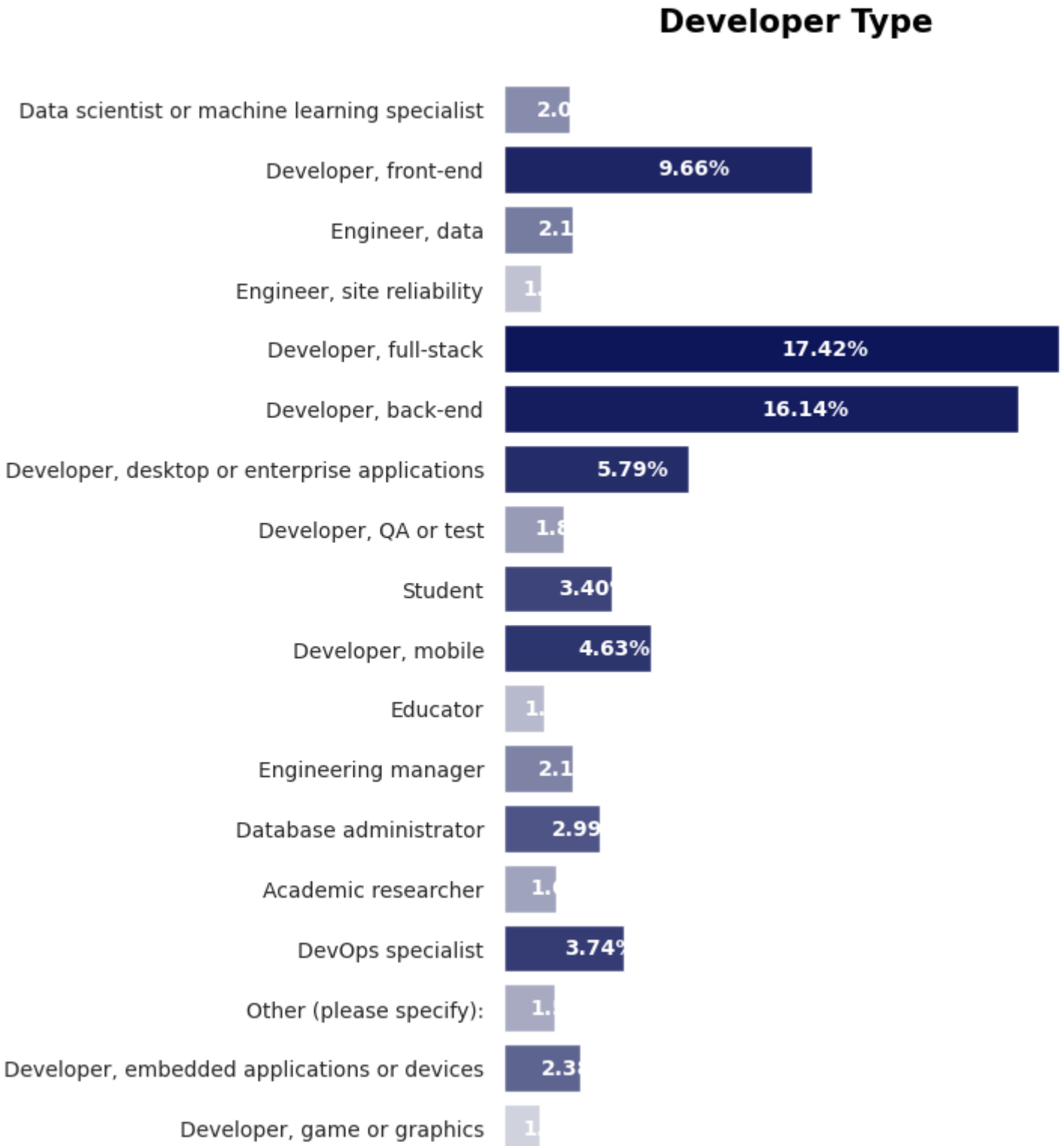
Out[72]:
```
'Which of the following describes your current job? Please select all that apply.'
```

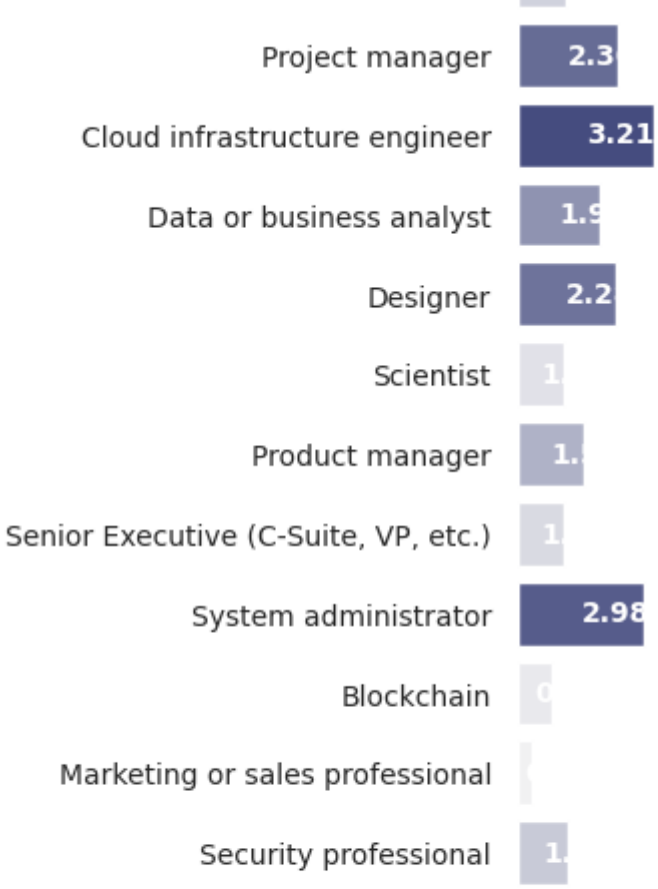
In [73]:

```
1 dt = colum_expand(survey_df.DevType)
```

In [74]:

```
1 custom_plot(dt, color='light:#000C66', y_label_font_size=10, percent_font_size=10, title='Developer Type')
```


Developer Type
Data scientist or machine learning specialist
2.0
Developer, front-end
9.66%
Engineer, data
2.1
Engineer, site reliability
1.
Developer, full-stack
17.42%
Developer, back-end
16.14%
Developer, desktop or enterprise applications
5.79%
Developer, QA or test
1.8
Student
3.40
Developer, mobile
4.63%
Educator
1.
Engineering manager
2.1
Database administrator
2.99
Academic researcher
1.
DevOps specialist
3.74
Other (please specify):
1.
Developer, embedded applications or devices
2.3
Developer, game or graphics
1



Total Responses: 164790

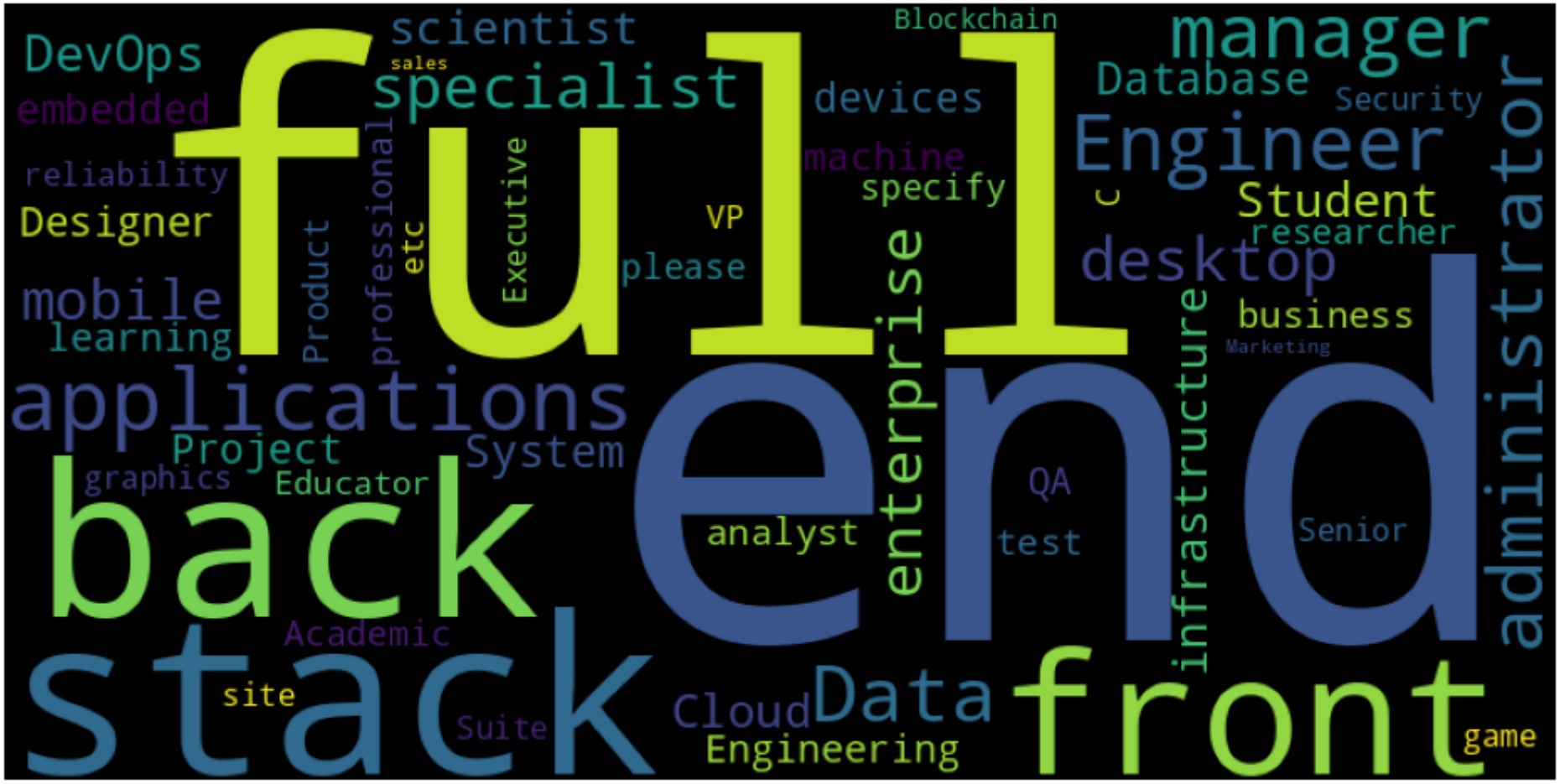
WordCloud for developer type

In [132]:
```
1 words = ' '.join((job for job in survey_df.DevType.dropna().str.replace(';', ' ').str.replace(',', ' ')))
```

In [133]:
```
1 words = words.replace('Developer', ' ')
```

In [80]:

```
# Generate word cloud
wordcloud = WordCloud(collocation_threshold=int(1e6), width=800, height=400,
                      background_color='black').generate(words)

# Plot the word cloud
plt.figure(figsize=(12, 6))
plt.imshow(wordcloud, )      # interpolation='bilinear'
plt.axis('off')
plt.show()
```



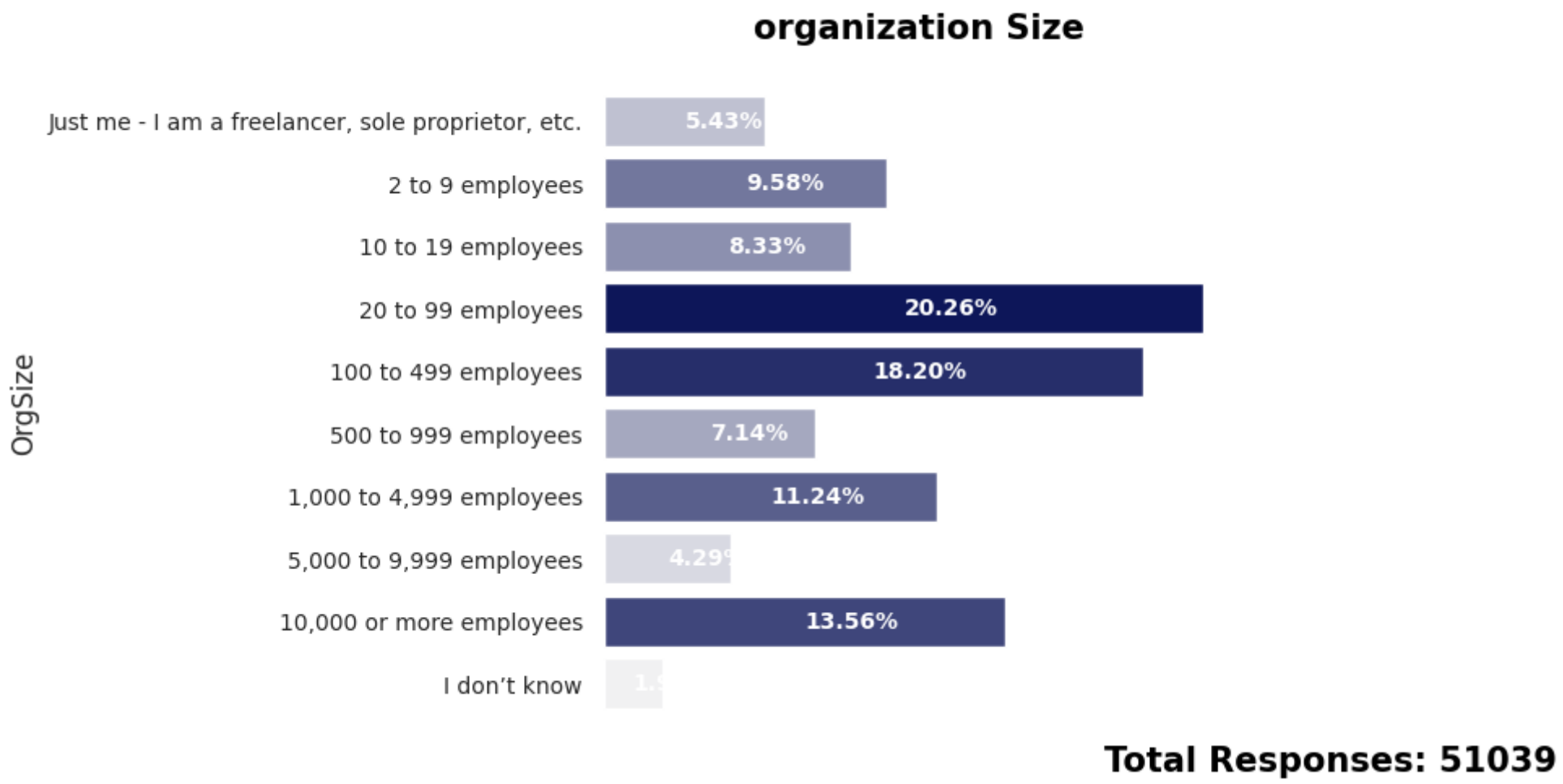
what is organization size of the developer...?

In [82]:
```
schema_df.OrgSize
```

Out[82]:
```
'Approximately how many people are employed by the company or organization you currently work for? '
```

In [83]:
```
reorder_list = [
    "Just me - I am a freelancer, sole proprietor, etc.",
    "2 to 9 employees", "10 to 19 employees", "20 to 99 employees",
    "100 to 499 employees", "500 to 999 employees",
    "1,000 to 4,999 employees", "5,000 to 9,999 employees",
    "10,000 or more employees", "I don’t know"
]

os = survey_df.OrgSize.value_counts().reindex( reorder_list )
```

In [84]:
```
custom_plot( os, plot_height=5, color='light:#000C66' , title="organization Size",
            percent_font_size=10, y_label_font_size=10)
```



donut plot function using pie plot

In [91]:
```
def plot_pie(data , title='', distance_btwn_pieces=0.09, startangle=-11):
    explode = (distance_btwn_pieces,) * len(data)
    plt.figure(figsize=(14,10))

    plt.pie( data, explode=explode,  labels=data.index, pctdistance=0.75,
             colors = ['red', 'blue', 'yellow','pink','blue'],
             wedgeprops={'linewidth': 1.5, 'edgecolor' : "green" },
             textprops={"weight":'bold', "size":20, 'family':'serif'},
             autopct='%1.1f%%',  startangle=startangle, shadow=True,
           )

    #plt.setp(pcts, color='black')
    hfont = {'fontname': 'serif', 'weight': 'bold'}
    plt.title(title, size=25, **hfont)

    centre_circle = plt.Circle((-0.08,0), 0.5, fc='white')
    fig = plt.gcf().gca().add_artist(centre_circle)
    ;
```

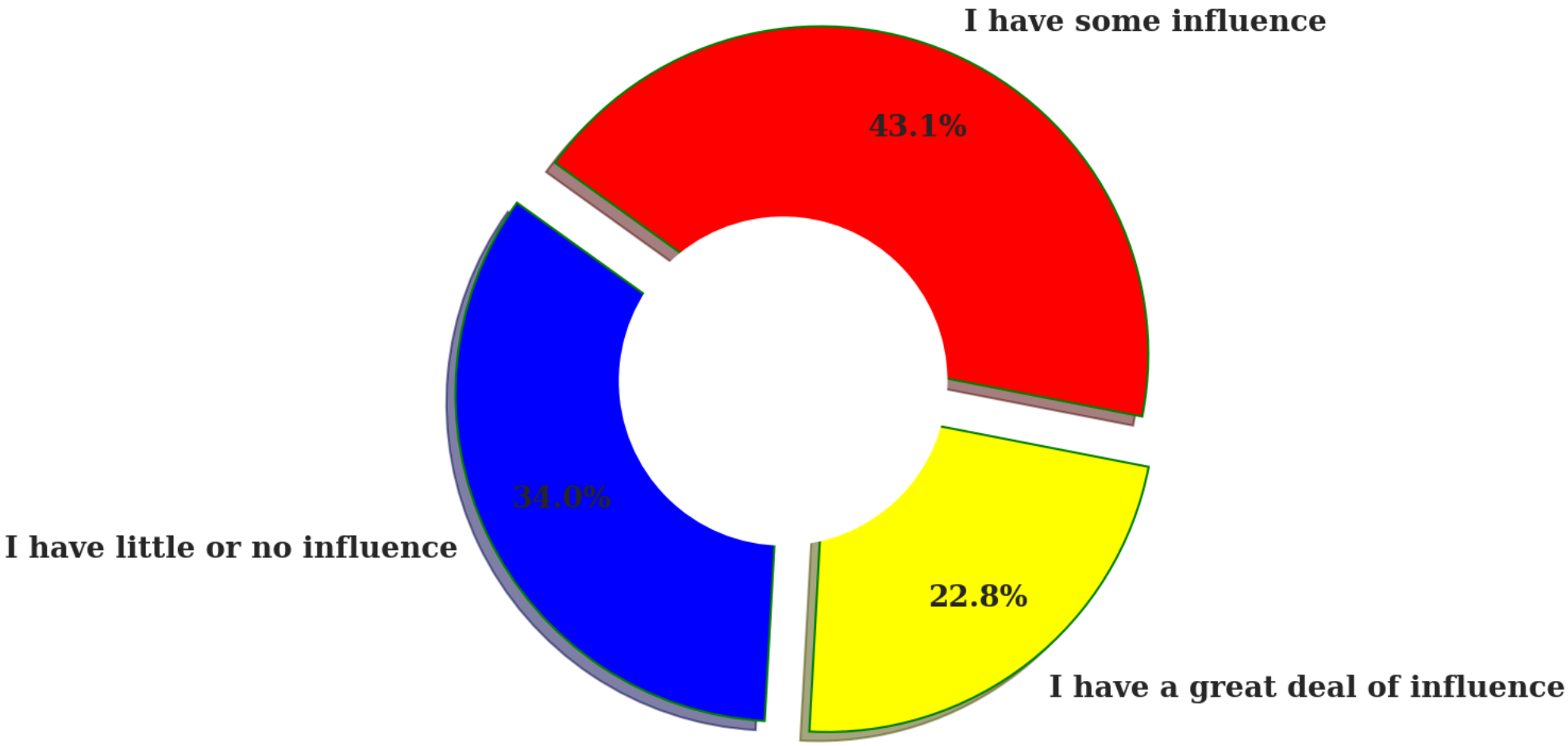
What level of influence developer, have over new technology purchases at your organization?

In [92]:
```
# PurchaseInfluence

schema_df.PurchaseInfluence
```

Out[92]:
```
'What level of influence do you, personally, have over new technology purchases at your organization?'
```

In [93]:
```
pi = survey_df.PurchaseInfluence.value_counts()
```

In [94]:
```
1 plot_pie(pi)
```



In [ ]:

```
def shorten_names(s):
    if s == 'I have some influence':
        return 'some influence'
    elif s == 'I have little or no influence':
        return 'little/no influence'
    elif s == 'I have a great deal of influence':
        return 'great influence'

tech_influence = survey_df.PurchaseInfluence.apply(shorten_names)
tech_influence = tech_influence.value_counts()

plot_pie(data=tech_influence,
         title = schema_df.PurchaseInfluence,
         distance_btwn_pieces=0.03
        )
```

Where do developer live?

In [95]:

```
# country

schema_df.Country
```

Out[95]:

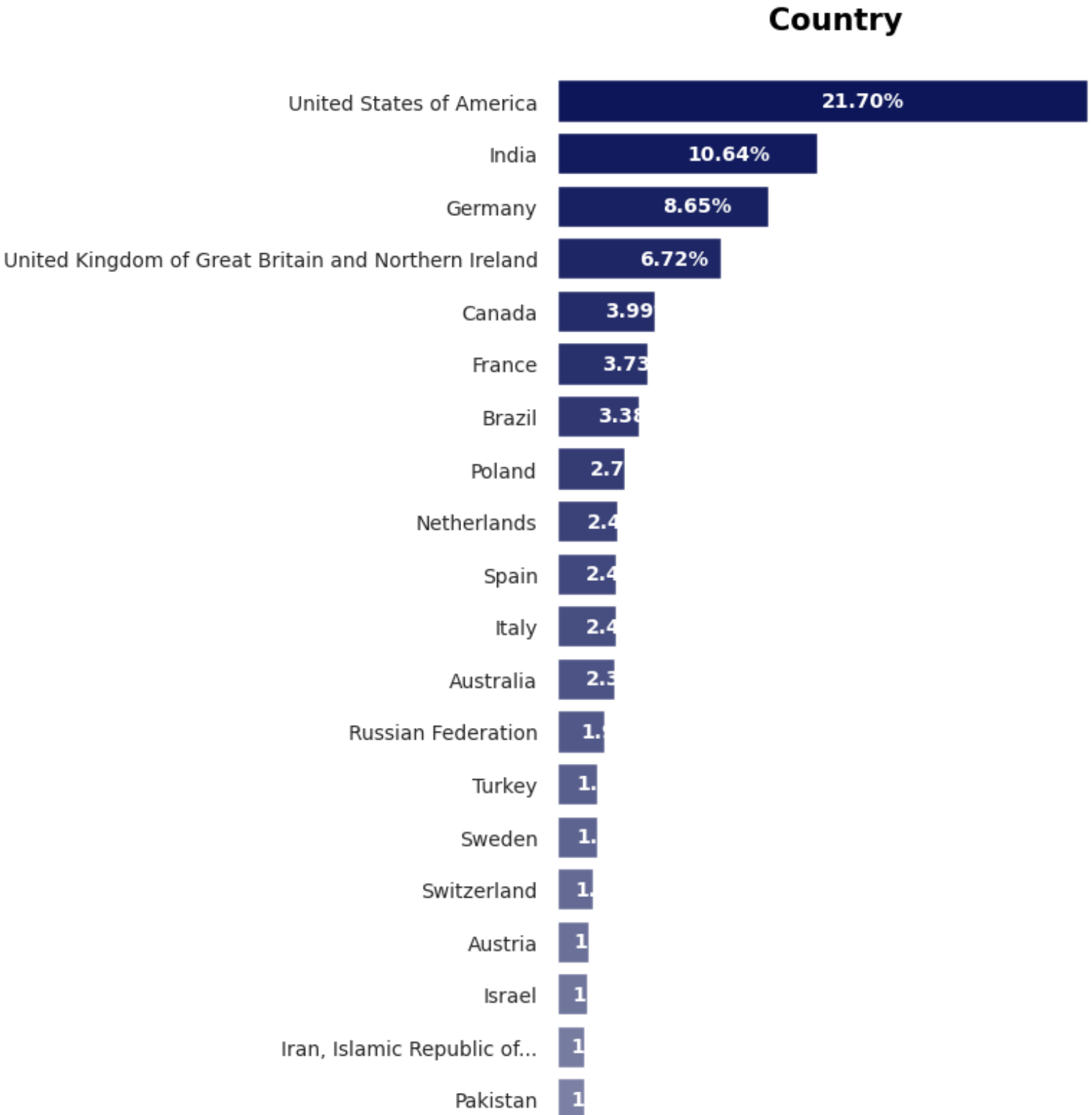
```
'Where do you live? <span style="font-weight: bolder;">*</span>'
```

In [96]:

```
c = survey_df.Country.value_counts()[:40]
```

In [97]:

```
custom_plot(c, color = 'light:#000C66', y_label_font_size=10, percent_font_size=10, title='Country',
            plot_height=19)
```


Country
United States of America
21.70%
India
10.64%
Germany
8.65%
United Kingdom of Great Britain and Northern Ireland
6.72%
Canada
3.99
France
3.73
Brazil
3.38
Poland
2.7
Netherlands
2.4
Spain
2.4
Italy
2.4
Australia
2.3
Russian Federation
1.
Turkey
1.
Sweden
1.
Switzerland
1.
Austria
1
Israel
1
Iran, Islamic Republic of...
1
Pakistan
1
try

Coun

Czech Republic 1

China 1

Belgium 1

Bangladesh 0

Ukraine 0

Romania 0

Mexico 0

Portugal 0

Greece 0

Denmark 0

Indonesia 0

Argentina 0

Nigeria 0

South Africa 0

Norway 0

Finland 0

Hungary 0

New Zealand 0

Egypt 0

Philippines 0

Total Responses: 62397

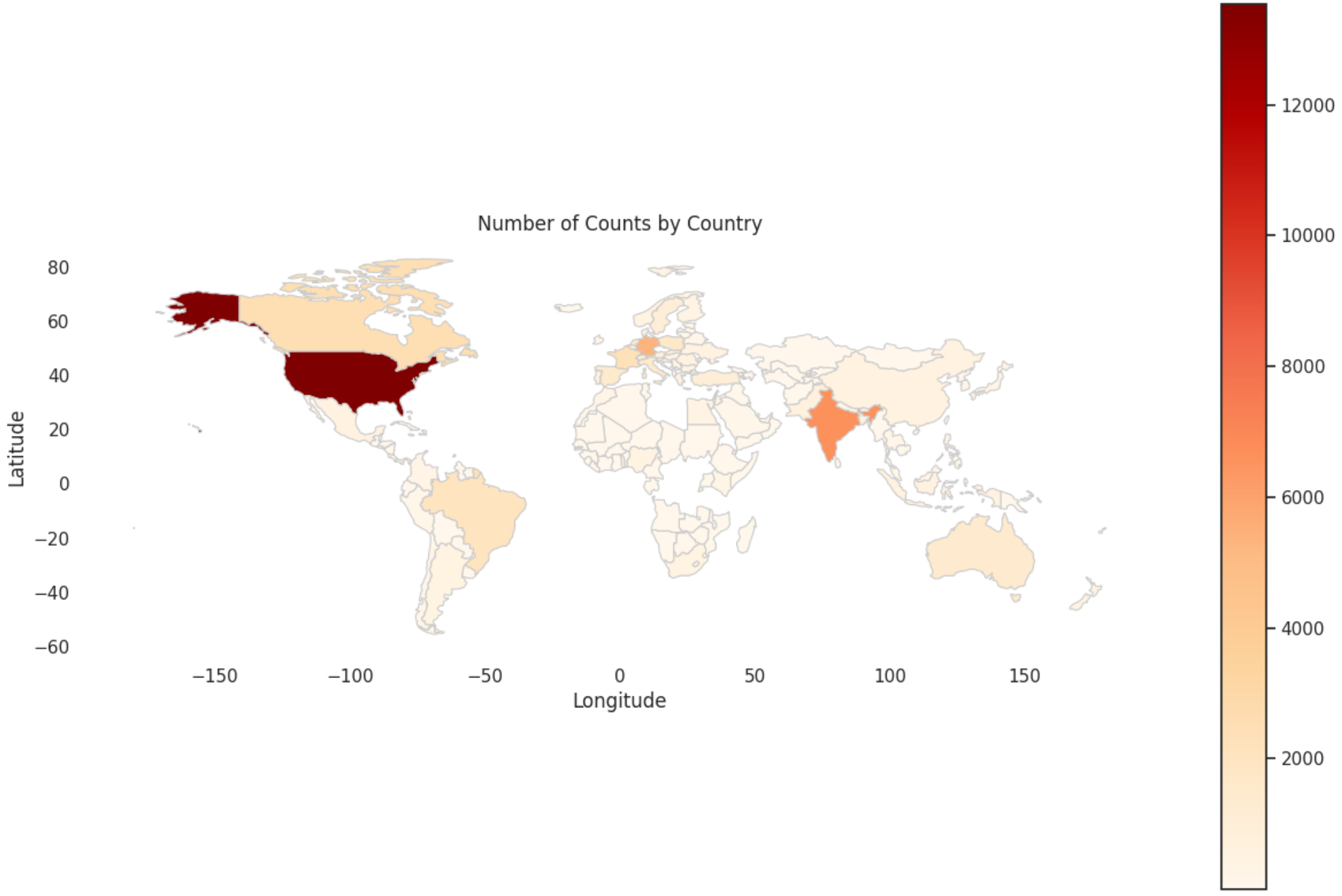
Map plot of country with developer count

In [98]:
```
1 d = survey_df.Country.value_counts().reset_index()
```

In [99]:
```
1 d.to_csv('country_for_map.csv')
```

In [100]:

```
import geopandas as gpd

data = pd.read_csv("country_for_map.csv")

# Load the world map
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

# Merge the world map with the data DataFrame
world = world.merge(data, how='left', left_on='name', right_on='Country')

# Plot the map
fig, ax = plt.subplots(1, 1, figsize=(15, 10))
world.plot(column='count', cmap='OrRd', linewidth=0.8, ax=ax, edgecolor='0.8', legend=True)

plt.title('Number of Counts by Country')
plt.xlabel('Longitude')
plt.ylabel('Latitude')

plt.show()
```



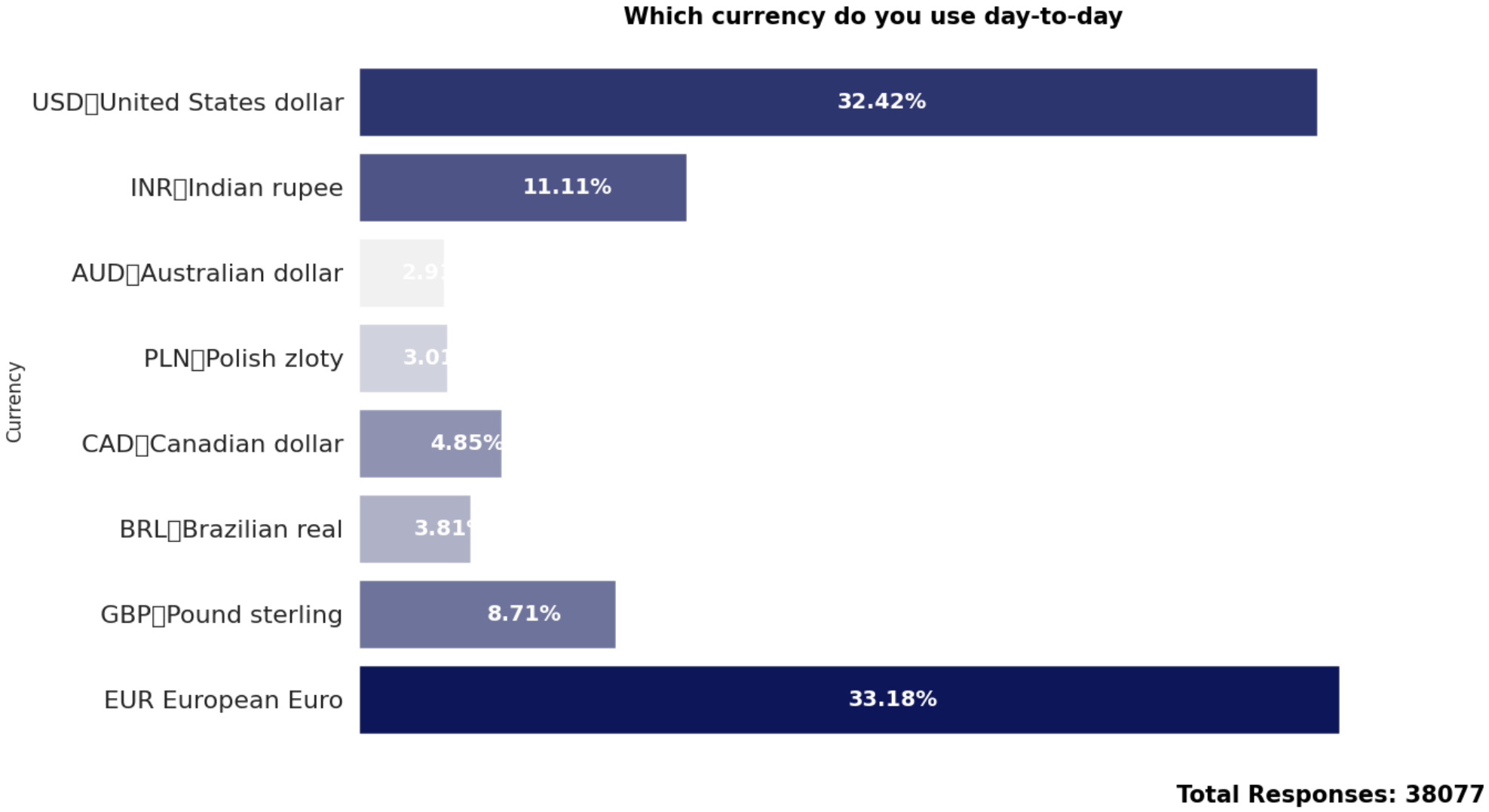
Which currency does developer use day-to-day?

In [101]:
```
1 survey_df.Currency.value_counts()
```

Out[101]:
```
Currency
EUR European Euro              12634
USD\tUnited States dollar      12346
INR\tIndian rupee               4229
GBP\tPound sterling             3318
CAD\tCanadian dollar            1847
                               ...
BND\tBrunei dollar                 1
PGK\tPapua New Guinean kina        1
SHP\tSaint Helena pound            1
GIP\tGibraltar pound               1
TOP\tTongan pa’anga                1
Name: count, Length: 142, dtype: int64
```

In [102]:

```
currency = survey_df.Currency.value_counts()[:8]
currency = currency.sample(len(currency))

custom_plot(currency, plot_height=8, plot_width=12,
            color = 'light:#000C66',
            title=schema_df.Currency.split('?')[0],
            y_label_font_size=16)
```



VersionControlSystem

In [103]:
```
# VersionControlSystem
```

In [104]:
```
schema_df.VersionControlSystem
```

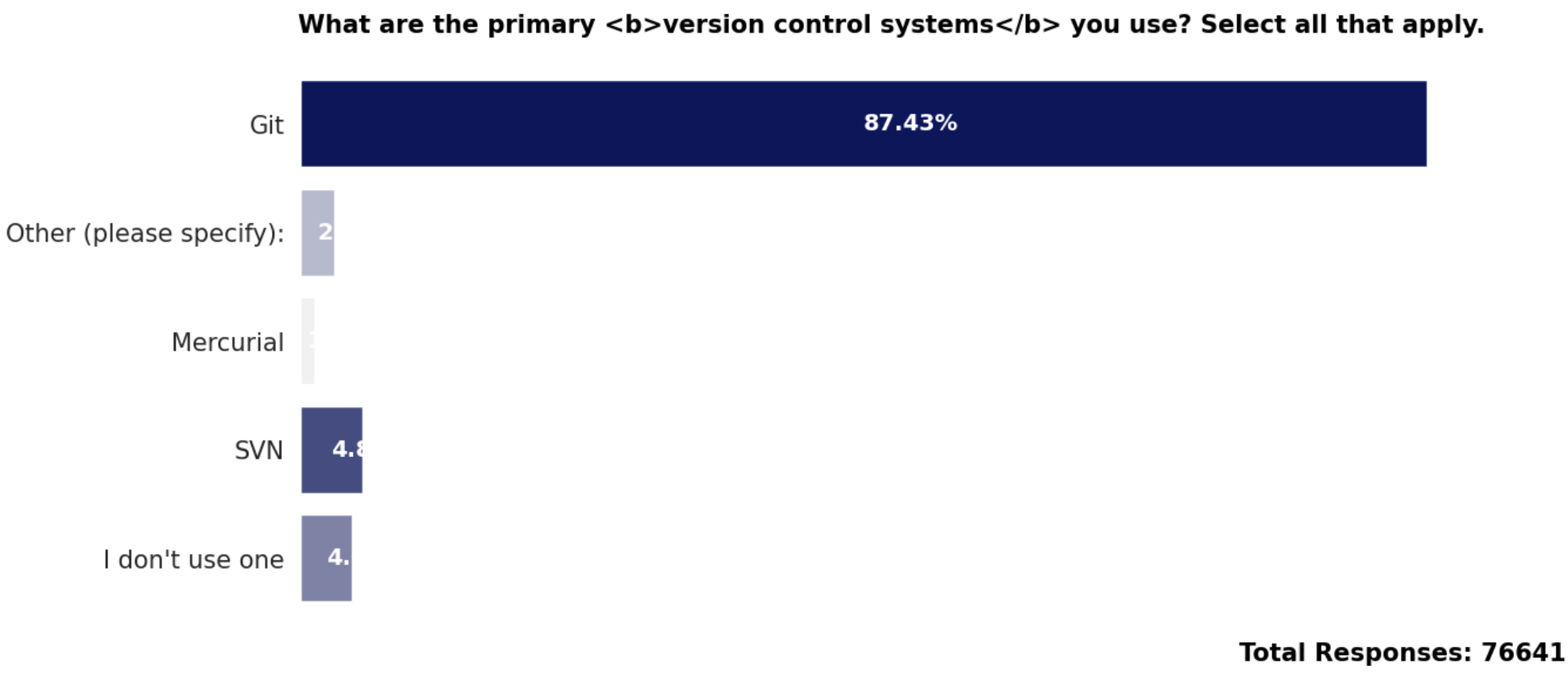
Out[104]:
```
'What are the primary <b>version control systems</b> you use? Select all that apply.'
```

In [105]:
```
colum_expand(survey_df.VersionControlSystem)
```

Out[105]:
```
Git                          67006
Other (please specify):       2047
Mercurial                      808
SVN                           3700
I don't use one               3080
dtype: int64
```

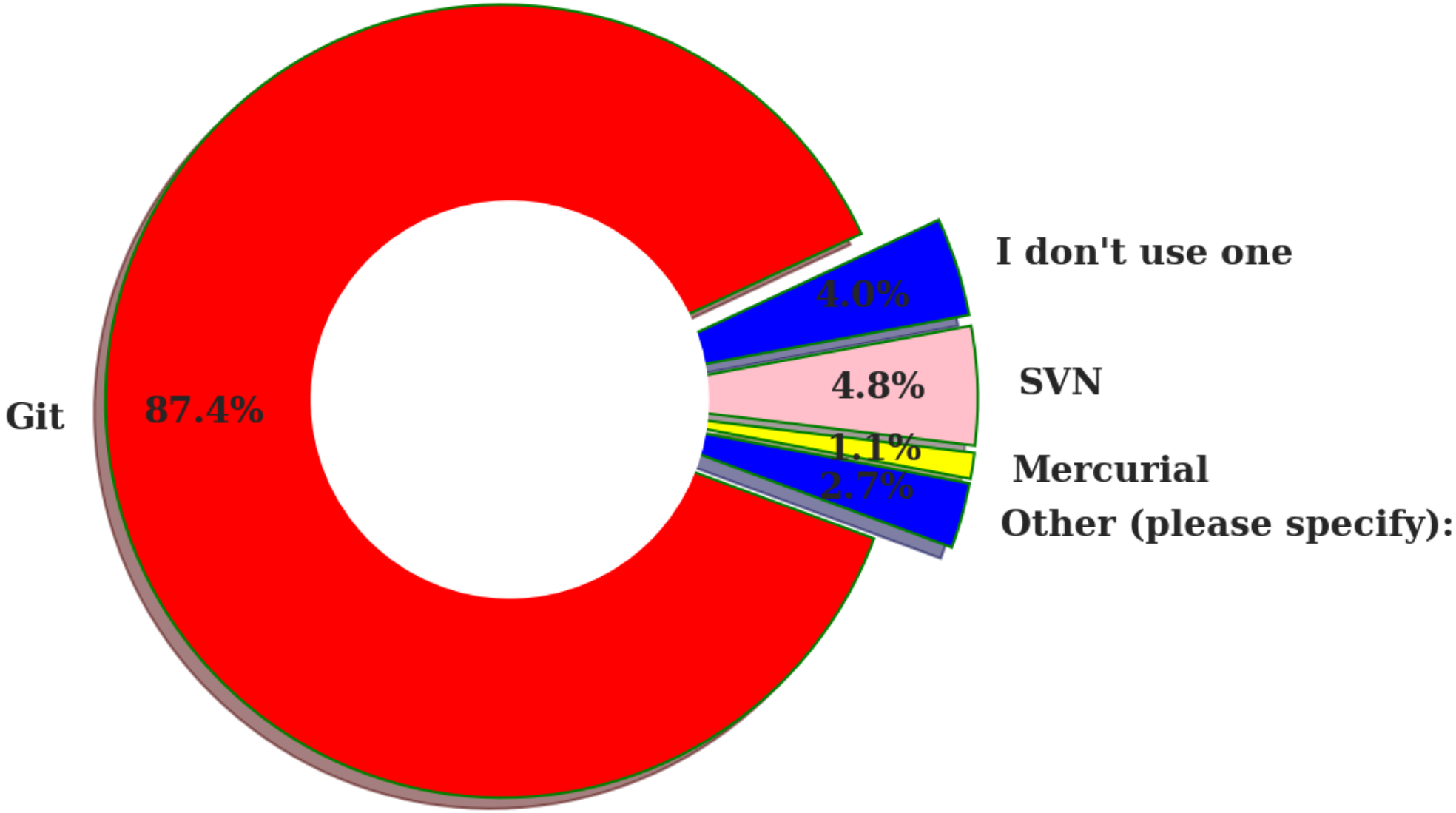
In [106]:

```
vcs = colum_expand(survey_df.VersionControlSystem)

custom_plot(vcs, plot_height=6, plot_width=13, color = 'light:#000C66',
            y_label_font_size=15, title=schema_df.VersionControlSystem)
```



In [107]:

```
plot_pie(vcs, startangle=25,
        distance_btwn_pieces=0.1)
```



what is your gender..?

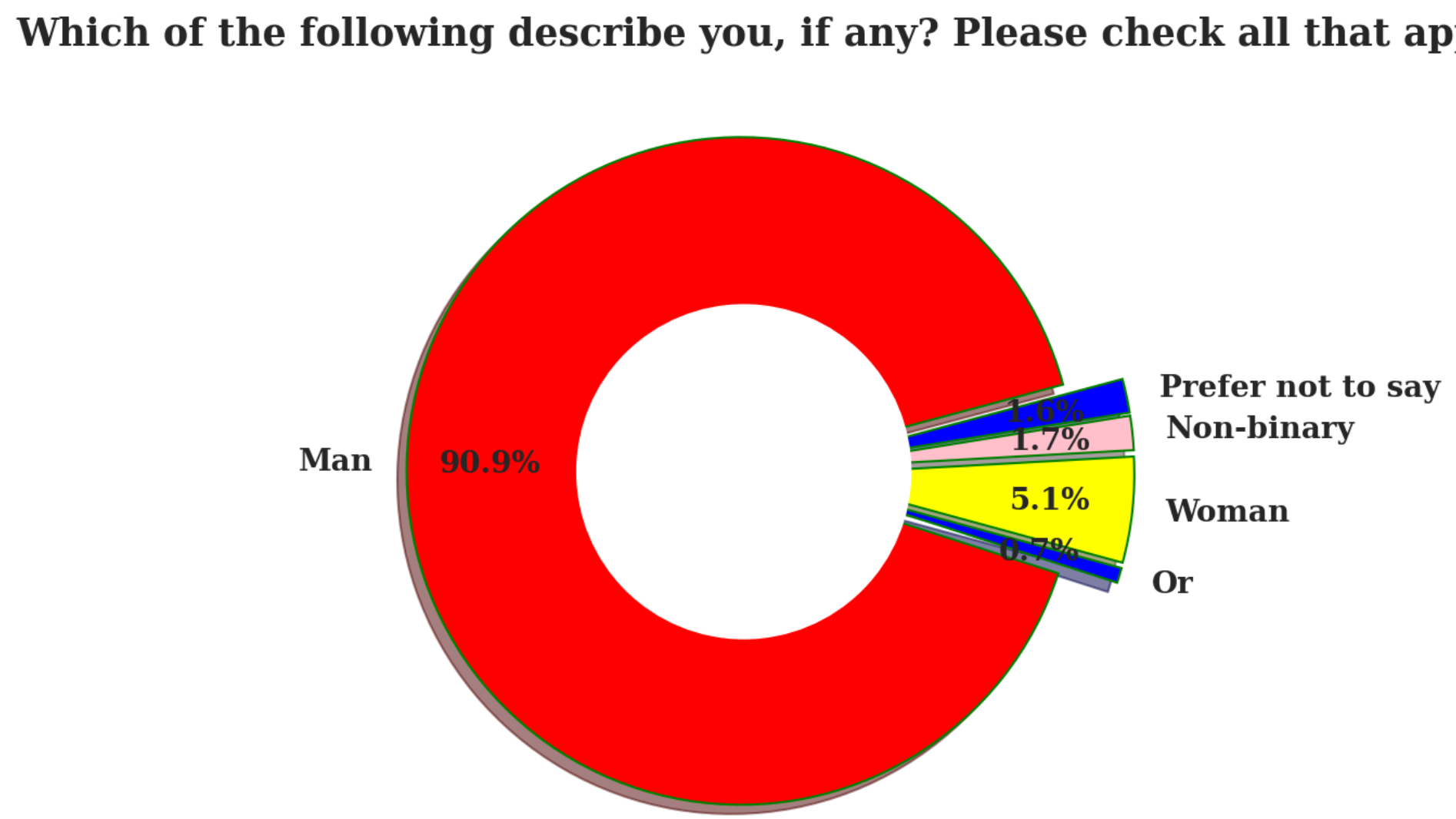
In [108]:
```
1 # Gender
```

In [109]:
```
1 colum_expand(survey_df.Gender)
```

Out[109]:
```
Man                                                   65097
Or, in your own words:                                  521
Woman                                                  3662
Non-binary, genderqueer, or gender non-conforming      1186
Prefer not to say                                      1172
dtype: int64
```

In [110]:

```
gender = colum_expand(survey_df.Gender)

gender.rename( lambda x: x.split(',')[0], inplace=True )

plot_pie(gender,distance_btwn_pieces=0.09, startangle=15,
        title=schema_df.Gender)
```



ethincity of developer

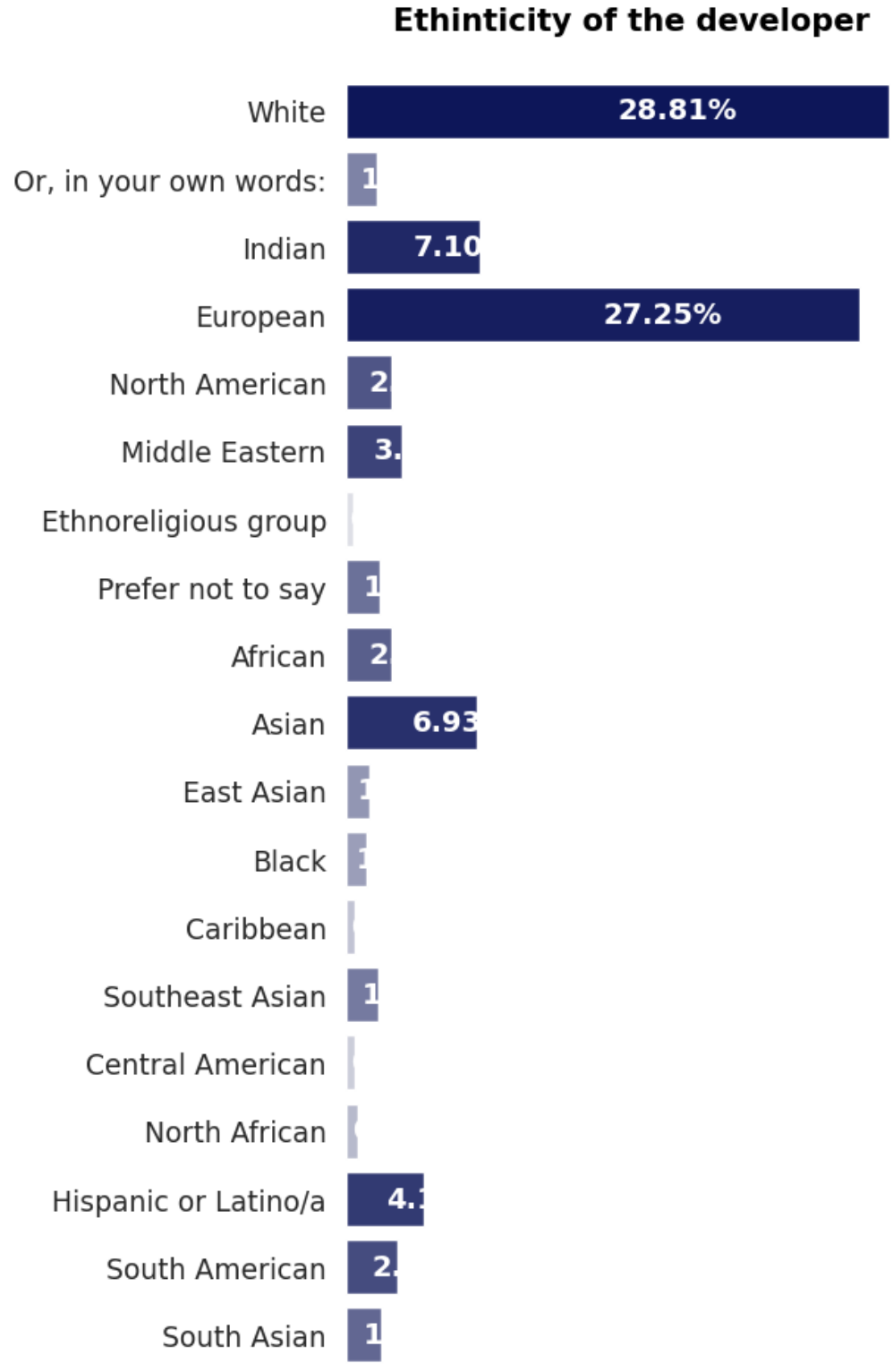
In [111]:
```
1 schema_df.Ethnicity
```

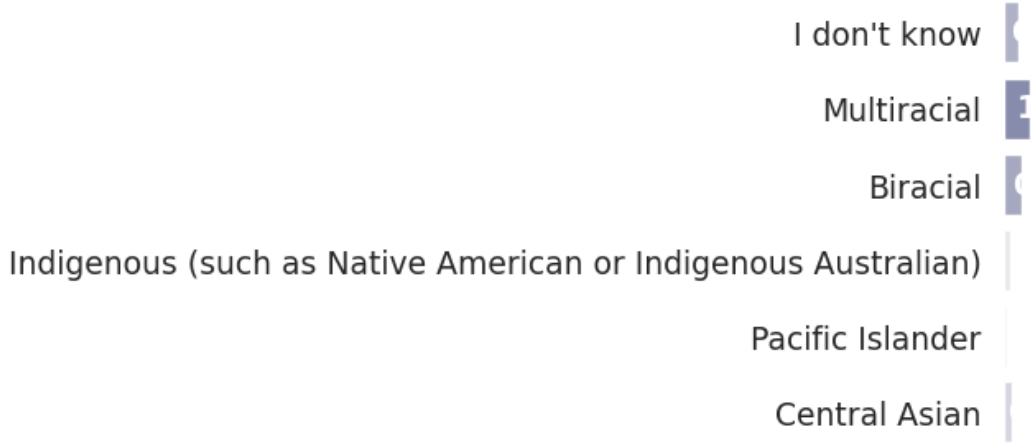
Out[111]:
```
'Which of the following describe you, if any? Please check all that apply.'
```

In [112]:
```
1 eth = colum_expand(survey_df.Ethnicity)
```

In [113]:

```
1 custom_plot( eth, color = 'light:#000C66' , title='Ethinticity of the developer')
```


Ethinticity of the developer
White
28.81%
Or, in your own words:
1
Indian
7.10
European
27.25%
North American
2
Middle Eastern
3.
Ethnoreligious group
Prefer not to say
1
African
2
Asian
6.93
East Asian
Black
Caribbean
Southeast Asian
1
Central American
North African
Hispanic or Latino/a
4.
South American
2
South Asian
1



Total Responses: 94971

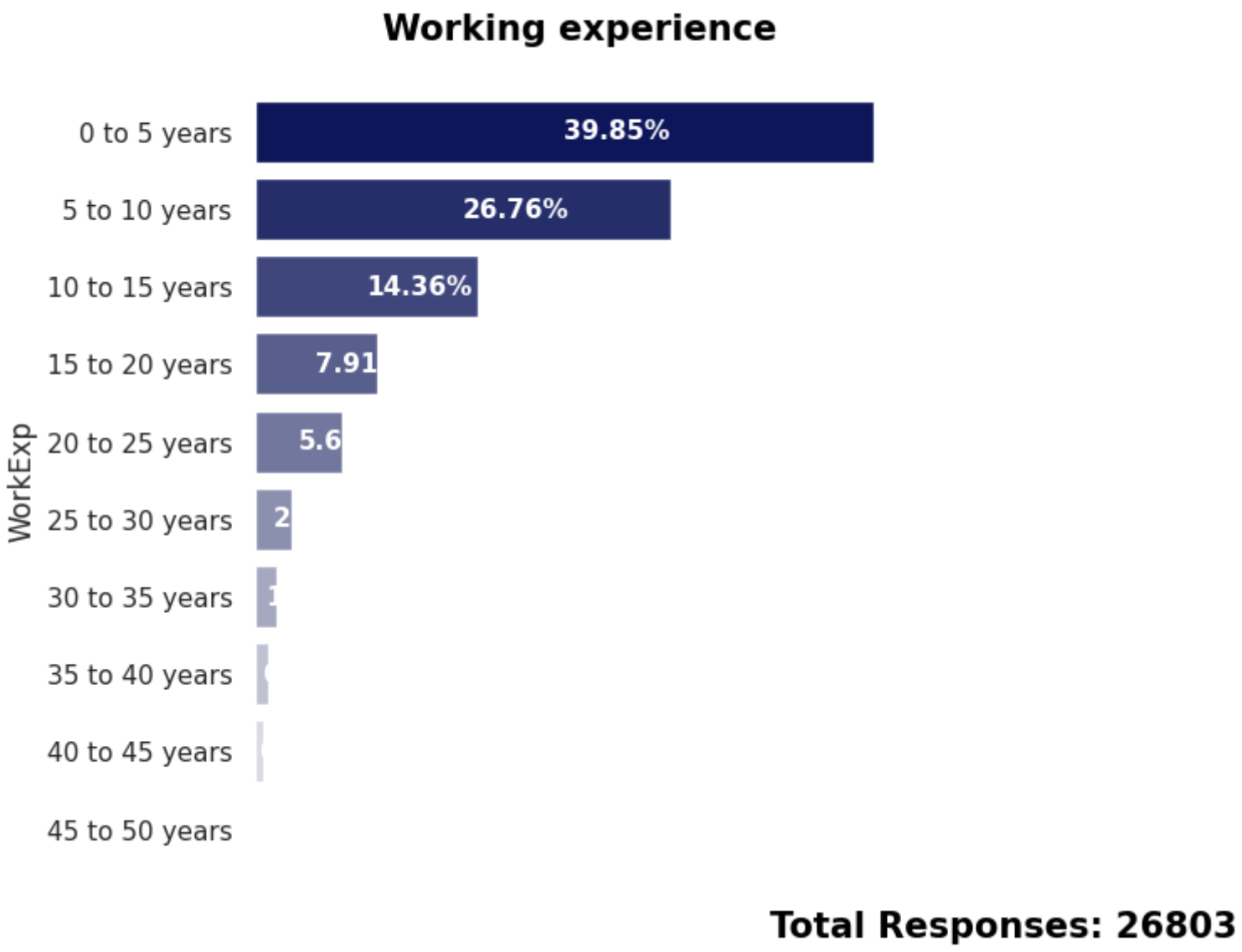
How many years of working experience do you have?

In [114]:
```
1 # WorkExp
2 schema_df.WorkExp
```

Out[114]: 'How many years of working experience do you have?'

In [115]:
```
1 we = survey_df.WorkExp.apply( make_groups ).value_counts()
```

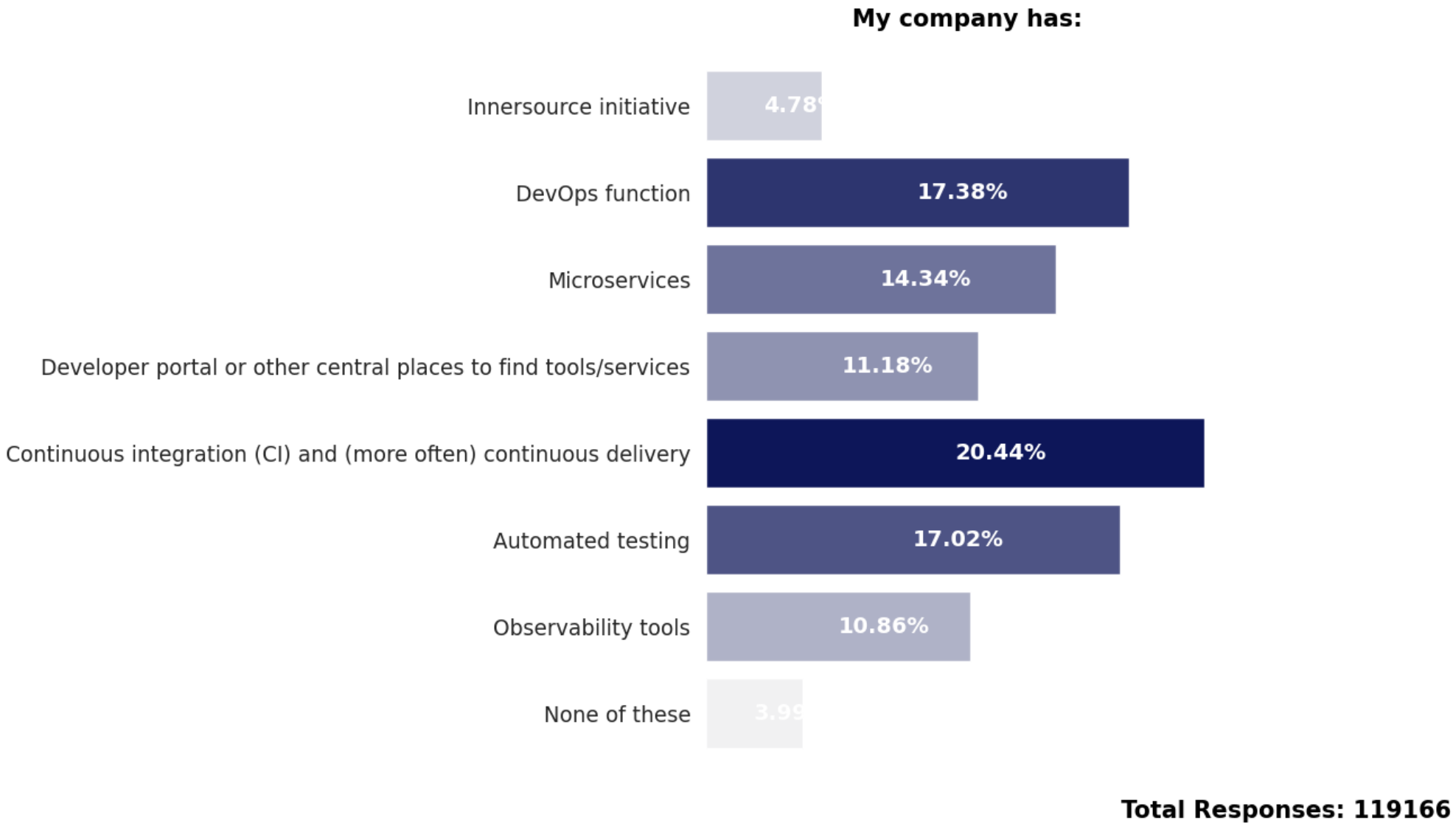
In [116]:
```
custom_plot(we, plot_height=6, color = 'light:#000C66', title='Working experience',
            y_label_font_size=11, percent_font_size=11 )
```



which technologies does your company have?

In [120]:

```
tech = colum_expand(survey_df.ProfessionalTech)

custom_plot(tech, plot_height=8, plot_width=6, color = 'light:#000C66',
            title=schema_df.ProfessionalTech)
```



Which programming, scripting, and markup languages have you done extensive development

work in over the past year

In [121]:
```
# LanguageHaveWorkedWith
```

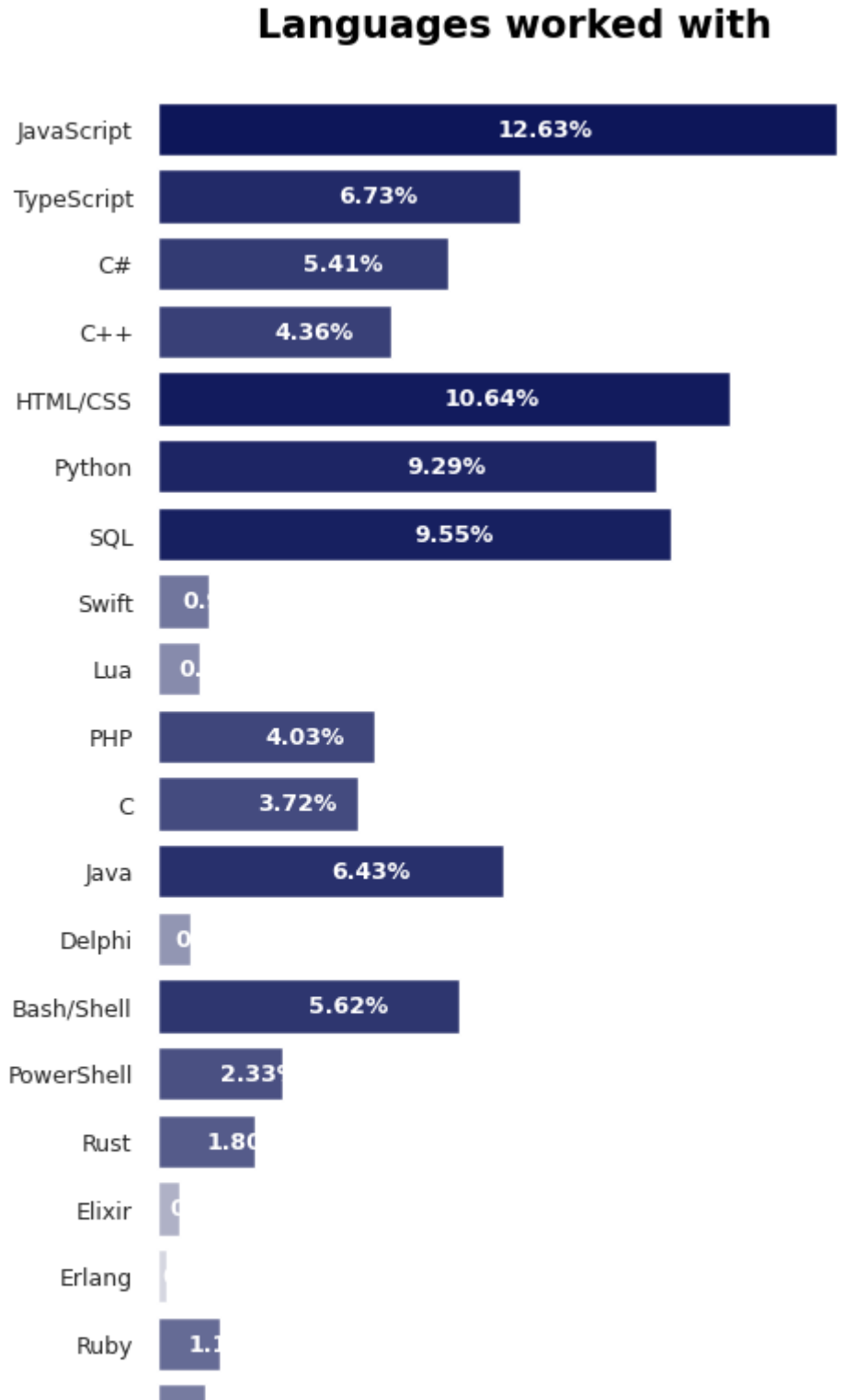
In [122]:
```
schema_df.Language
```

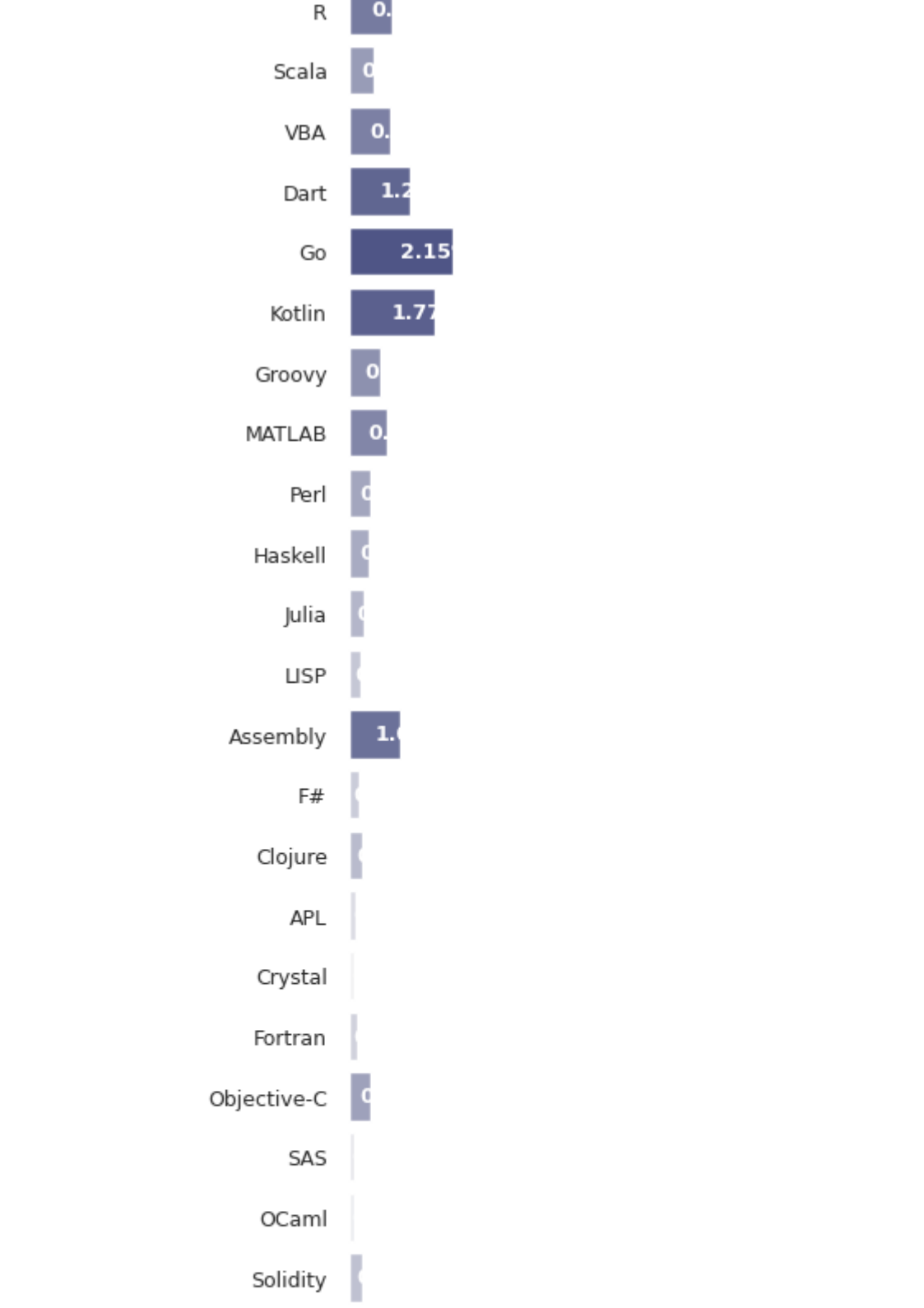
Out[122]:
```
'Which <b>programming, scripting, and markup languages</b> have you done extensive development work in over the past y
ear, and which do you want to work in over the next year? (If you both worked with the language and want to continue t
o do so, please check both boxes in that row.)'
```

In [123]:
```
lhw = colum_expand(survey_df.LanguageHaveWorkedWith)
```

In [124]:

```
custom_plot( lhw, color = 'light:#000C66', title='Languages worked with',
            percent_font_size=9, y_label_font_size=9, plot_height=20)
```


Languages worked with
JavaScript
12.63%
TypeScript
6.73%
C#
5.41%
C++
4.36%
HTML/CSS
10.64%
Python
9.29%
SQL
9.55%
Swift
Lua
PHP
4.03%
C
3.72%
Java
6.43%
Delphi
Bash/Shell
5.62%
PowerShell
Rust
Elixir
Erlang
Ruby


R
Scala
VBA
Dart
Go
Kotlin
Groovy
MATLAB
Perl
Haskell
Julia
LISP
Assembly
F#
Clojure
APL
Crystal
Fortran
Objective-C
SAS
OCaml
Solidity

COBOL

Total Responses: 367821

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:
```
1
```

In [ ]:

```
survey_df.WebframeHaveWorkedWith      # django flask FastAPI, Node JS
survey_df.WebframeWantToWorkWith

survey_df.LanguageWantToWorkWith

survey_df.DatabaseHaveWorkedWith
survey_df.DatabaseWantToWorkWith

survey_df.PlatformHaveWorkedWith
survey_df.PlatformWantToWorkWith

survey_df.MiscTechHaveWorkedWith
survey_df.MiscTechWantToWorkWith

survey_df.ToolsTechHaveWorkedWith
survey_df.ToolsTechWantToWorkWith

survey_df.CompTotal                   # annual income
survey_df['OpSysPersonal use']        # operating system
```

E N D