Data Bootcamp Final Project: Accounting for US Healthcare Spending Growth From 1960 to 2015

Author: **Ziang (Leon) Wang** | Last Updated: **December 21, 2017**

Feel free to contact me via the following methods:

Email: ziang.wang@stern.nyu.edu (mailto: ziang.wang@stern.nyu.edu) | **LinkedIn:** My Linkedin Profile (https://www.linkedin.com/in/ziangwang/)

Background and Goal

Over the past few decades, growth in US per-capita healthcare expenditures has outpaced inflation, and it has grown from 5% of GDP in 1960 to 17.8%, as of 2015 (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/Tables.zip). Consequently, controlling cost growth is a priority for all stakeholders, political and otherwise. Often, decision-makers focus on specific initiatives, and few have objectively demonstrated a comprehensive understanding of the health expenditures growth.

Thus, this project aims to re-examine the Centers for Medicare & Medicaid Services (CMS) data and narrate what has happened to the spending growth across the entire healthcare system at the national level. With this information, leaders can better prioritize their agenda.

Key Questions

The overarching objectives is: how has health expenditures grown over the past two decades (1995 to 2015)?

Specifically:

- **Sources:** Which payors have faced the most significant in per capita spending growth?
- **Uses:** What product/service segments have experienced the most per-capita spending growth?

These analyses are important to provide as valuable historical context for the fiscal year 2016 report that CMS has recently published on a similar topic (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/PieChartSourcesExpenditures.pdf).

Agenda

1. **Data Report:** description of NHEA data and accessing/sourcing from CMS
2. **Set-up:** discussion of tools and packages I used to analyze the data & setup this notebook
3. **Organization:** input, cleaning, and manipulation of the dataset to get ready for subsequent analyses
4. **Analysis:** graphical analyses of key questions and related commentary
5. **Discussion:** implications and next steps to this analysis

1. Data Report

Access: This project focuses on the National Health Expenditure Accounts (NHEA) dataset, kindly provided by CMS through their official website for historical data (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html). Access is free and does not require any applications for access. The aggregate, national-level historical data is available as .csv and .xlsx (excel) files under the link for "National Health Expenditures by type of service and source of funds, CY 1960-2016 [ZIP, 210KB]". I focused on this primary dataset because it provides a lot of rich details and interesting insights that are relevant to the central, driving questions of this project. However, I also discuss related datasets provide on the same website that get at other interesting questions I did not get a chance to explore during this project.

Content: These are the official estimates of total health care spending in the United States. Going all the way back to 1960, the NHEA measures annual U.S. expenditures for health care goods and services, public health activities, government administration, the net cost of health insurance, and investment related to health care. The historical data, spanning 1960 to 2015, is presented by type of service and sources of funding. These categories are explained throughout *4. Analysis*. What is great about this data is that it's multi-dimensional, MECE (mutually exclusive, comprehensive exhaustive), and consistent over time.

Format: These files were originally formatted to resemble "financial statements" that "account" for national health expenditures. As a result, there will be indents/spaces to deal with." I recommend inputting the Excel file because the packages I used were able to easily take in the data and understand the NaN values, but they could not in the case of the .csv files.

Below is the detailed description of the CMS's methods according to the 2016 National Health Expenditure Accounts Methodology Paper. (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/dsm-16.pdf)

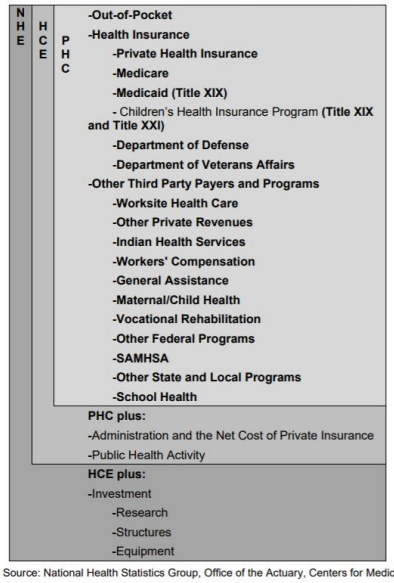
Expenditures in the NHEA represent aggregate health care spending in the U S. The NHEA recognize several types of health care spending within this broad aggregate.

- Personal Health Care expenditures (PHC) measures the total amount spent to treat individuals with specific medical conditions.
- Health Consumption Expenditures (HCE) represents spending for all medical care rendered during the year, and is the sum of PHC, government public health activity, and government administration and the net cost of health insurance.
- National Health Expenditures (NHE) equals HCE plus Investment, or the sum of medical sector purchases of structures and equipment and expenditures for noncommercial medical research.
- Government public health activity measures spending by governments to organize and deliver health services and to prevent or control health problems.
- Government administration and the net cost of health insurance includes the administrative cost of running various government health care programs, and the difference between premiums earned by insurers and the claims or losses incurred for which insurers become liable (the net cost of PHI).
- Finally, the category of Investment includes spending for noncommercial biomedical research and expenditures by health care establishments on structures and equipment.

All graphs below are from the same paper. (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/dsm-16.pdf)

Below is an organization chart of the **different payor categories** variables included and how they are related.

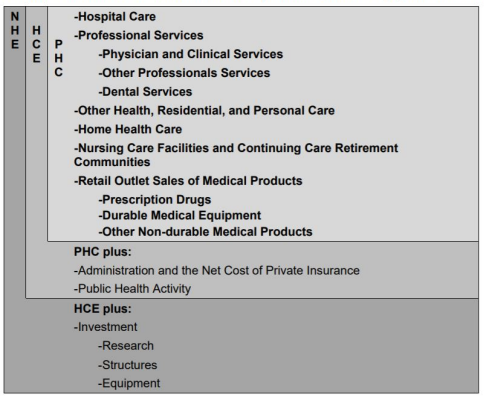
Exhibit 5: Structure of the National Health Expenditure Accounts by Source of Funds



Source: National Health Statistics Group, Office of the Actuary, Centers for Medicare & Medicaid Services.

Below is an organization chart of the **different spending categories** variables included and how they are related.

Exhibit 3: Structure of the National Health Expenditure Accounts by goods and services



Source: National Health Statistics Group, Office of the Actuary, Centers for Medicare & Medicaid Services.

2. Set-Up

In the analyses detailed below, I used the following Python packages:

- *pandas* package: core tool for data import, manipualtion, merging, and analysis.
- *matplotlib* pakcage: foundation of plots
- *datetime* package: allows for easy references to dates
- *numpy* package: math operations and transformations
- *requests & io* packages: these are tools that allow for easy download or inputs of zipped files
- *zipfile* package: allows for conversion of zipped files to bytes and readible formats

I also used some additional packages that we did not learn in class:

- *plotly* packages: plots functions easily interpreting the data, based on matplotlib methods
- *cufflinks* package: gives plotly-based plots their signature look

In [1]:
```
import pandas as pd                          # data package
import matplotlib as mpl                     # graphics package
import matplotlib.pyplot as plt              # pyplot module
import datetime as dt                        # date and time module
import numpy as np                           # foundation for pandas math operations
import requests, io                          # internet and input tools
import zipfile as zf                         # zip file tools
from plotly.offline import iplot, iplot_mpl  # plotting functions
import plotly.graph_objs as go               # ditto
import plotly                                # just to print version and init notebook
import cufflinks as cf                       # gives us df.iplot that feels like df.plot
cf.set_config_file(offline=True, offline_show_link=False)

# these lines make graphics show up in the notebook
%matplotlib inline
plotly.offline.init_notebook_mode(connected=True)

# check versions of major packages
print('Pandas version: ', pd.__version__)
print('Matplotlib version: ', mpl.__version__)
print('Plotly version: ', plotly.__version__)
```

```
Pandas version:  0.21.0
Matplotlib version:  2.0.2
Plotly version:  2.2.1
```

3. Organization

Below are the steps to input, clean, and manipulate the national-level, historical NHEA data. It includes both payor (who paid) and services (what was it spent on) information. I made new dataframes that included on the datasets I needed for subsequent analyses.

Here is the hyperlink to the dataset I focus on:

In [2]:
```
#National Health Expenditure Accounts Data
url1a = 'https://www.cms.gov/Research-Statistics-Data-and-Systems/'
url1b = 'Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/NHE2015.zip'
url1 = url1a + url1b #Location of the National Healthcare Expenditure Data
```

Download and convert data to readible format:

In [3]:
```
r1 = requests.get(url1) # get "response" from url
nhea = zf.ZipFile(io.BytesIO(r1.content)) # convert bytes to zip file
nhea.namelist() # here's what the zipped folder contains
```

Out[3]:
```
['NHE2015.csv', 'NHE2015.xls']
```

Key considerations about this dataset (I'm using the excel file):

- First row restates title of the file
- The next 37 rows are total spending in by each payor group
- The rest of the data are specific product/service segments of spending, broken down by payor (**due to this quirk of having multiple payors, I will load in the data two times to avoid this issue**)
- Last 3 rows are comments
- Thousands are separated by commas
- All values' units are set to millions of US dollars
- The second file in this folder is of interest (excel format) Below, I include a few rows of this data to give you a sense of the original data.

a. Payor Data

Read in payor data into a dataframe called "pay" & simultaneously cleaning it up, addressing commas, N/A values, and indexing.

In [4]:
```
# extract and read excel all national health expenditure accounts data
pay = pd.read_excel(
                    nhea.open(nhea.namelist()[1]), # 2nd file of the 2-file zipped folder is the excel version (index start = 0)
                    skiprows = 1, #first row is the title of the file, so it's irrelevant
                    skip_footer = 505, # Skip last 505 rows, which are detailed accounts of
                    index_col=0, # column index is set to the first column (python indexes starting from 0)
                    thousands = ',', # thousands separator is comma
                    na_values=[' - ', ' -', '- ', '-', 0.0], # define NaN values
                    )
pay.head()
```

Out[4]:

	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	2006	2007	2008	2009	2010	2011	
Expenditure Amount (Millions)																		
Total National Health Expenditures	27214	29138	31842	34595	38394	41852	46081	51565	58402	65923	...	2156516	2295650	2399061	2494669	2596405	2687902	27
Out of pocket	12949	13357	14255	15311	16928	18209	18594	18537	20500	22601	...	273196	289916	294881	293130	298727	308519	31
Health Insurance	7497	8236	8999	9892	10971	12023	15673	21109	24379	27565	...	1517815	1609716	1696163	1796114	1875102	1948166	20
Private Health Insurance	5812	6468	7178	7952	9052	10072	10296	10452	11830	13363	...	737559	776628	803020	832551	863064	895098	92
Medicare	0	0	0	0	0	0	1842	4924	6218	7045	...	403690	432751	466971	498872	519253	546273	56

5 rows × 56 columns

Variables of interest, payor and spending categories, are by default the index, so the dataset is *tranposed* for easier manipulation of the data, later on.

In [5]:
```
pay = pay.T # Payor categories are the index, years are the "variables; obviously not conducive to the analysis
```

For the payor information, I create a variable that only includes data for the 5 major payor categories:

In [6]:
```
#Select the major categories to get a macro perspective of what is important and what is not
pay_maj_cat_list = ['Total National Health Expenditures','     Out of pocket', '     Health Insurance',
                    '     Other Third Party Payers and Programs', '     Public Health Activity', '     Investment', 'POPULATION']
pay_maj_cat = pay[pay_maj_cat_list]
```

For convenience's sake, I renamed and shortened the variable names

In [7]:
```
#Rename columns
pay_maj_cat.columns = ['TOTAL','OUT_OF_POCKET','INSURANCE','OTHER_THIRD_PARTIES','PUBLIC_HEALTH','INVESTMENT','POPULATION']
```

Health insurance

I go on to look at the health insurance category specifically and look at the individual categories.

In [8]:
```
#Select the relevant categories of health insurance
insurance_list = ['     Health Insurance', '          Private Health Insurance', '          Medicare',
                  '          Medicaid (Title XIX)',  '          CHIP (Title XIX and Title XXI)',
                  '          Department of Defense', '          Department of Veterans Affairs', 'POPULATION']
insurance = pay[insurance_list]
```

For convenience's sake, I renamed and shortened the variable names

In [9]:
```
#Rename columns/variable names
insurance.columns = ['TOTAL_INSURANCE','PRIVATE','MEDICARE', 'MEDICAID','CHIP','MILITARY','VETERANS','POPULATION']
```

b. Spending Data

Read in spending data into a dataframe called "spend" & simultaneously cleaning it up, addressing commas, N/A values, and indexing.

In [10]:
```
spend = pd.read_excel(
                    nhea.open(nhea.namelist()[1]),
                    skiprows = 1,
                    skip_footer = 3, # Skip last 3 tows, which are unuseful comments
                    index_col=0, # column index is the 1st column (python indexes starting from 0)
                    thousands = ',', # thousands separator is comma
                    na_values=[' - ', ' -', '- ', '-', 0.0], # space-dash-space is demarked as a NaN value
                    )
```

Variables of interest, payor and spending categories, are by default the index, so the dataset is *tranposed* for easier manipulation of the data, later on.

In [11]:
```
spend = spend.T # Payor categories are the index, years are the "variables; obviously not conducive to the analysis
```

For the spending information, I create a variable that only includes data for "total spending" on each spending segment:

In [12]:
```
segment_list = ['Total National Health Expenditures','POPULATION','Total Hospital Expenditures',
                'Total Physician and Clinical Expenditures','Total Dental Services Expenditures',
                'Total Other Professional Services Expenditures','Total Home Health Care Expenditures',
                'Other Non-Durable Medical Products Expenditures', 'Total Prescription Drug Expenditures',
                'Total Durable Medical Equipment Expenditures',
                'Total Nursing Care Facilities and Continuing Care Retirement Communities',
                'Total Other Health, Residential, and Personal Care Expenditures',
                'Total Administration and Total Net Cost of Health Insurance Expenditures', 'Public Health Activity',
                'Research', 'Total Structures and Equipment']
segments = spend[segment_list]
```

For convenience's sake, I renamed and shortened the variable names

In [13]:
```
#Rename columns/variable names
segments.columns = ['TOTAL','POPULATION','HOSPITAL','PHYSICIAN','DENTAL','SERVICES','HOME_CARE',
                    'PRODUCTS','DRUGS','EQUIPMENT','NURSING','OTHER','ADMIN','PUBLIC_HEALTH',
                    'RESEARCH','INFRASTRUCTURE']
```

4. Analysis

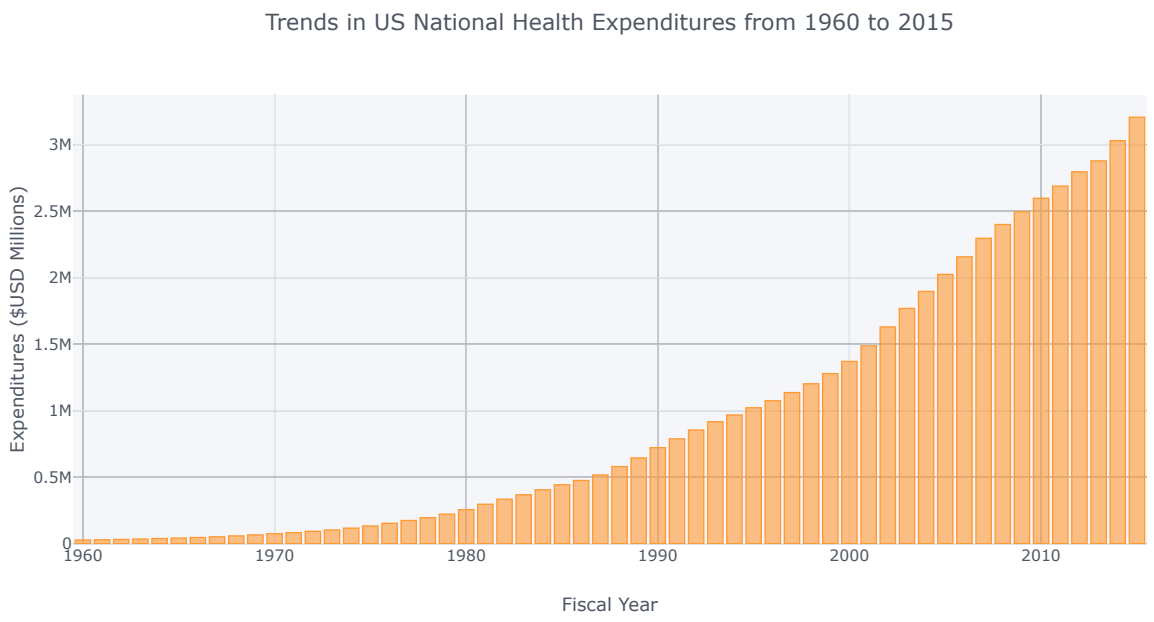
Overall growth in healthcare spending

It is interesting to first look at the growth of healthcare spending before moving forward.

In [14]:
```
# pull total spending
pay_maj_cat_total = pay_maj_cat['TOTAL']
```

The bar chart below shows how US national health expenditures have grown over time.

In [15]:
```
#Graph trends in Total National Health Expenditures
pay_maj_cat_total.iplot(
                        title='Trends in US National Health Expenditures from 1960 to 2015',
                        xTitle = 'Fiscal Year',
                        yTitle = 'Expenditures ($USD Millions)',
                        kind='bar',
                        )
```



In 1960, US spent 27 billion. In 2015, US spent over 3.2 trillion.

In [16]:
```
TOTAL_CAGR = (float(pay_maj_cat_total.iloc[-1])/float(pay_maj_cat_total.iloc[0]))**(1/(2015-1960))-1
print('Compounded annual growth rate (CAGR) of Total NHEs over 1960 to 2015 is:', round(TOTAL_CAGR*100,2),'%')
```

```
Compounded annual growth rate (CAGR) of Total NHEs over 1960 to 2015 is: 9.06 %
```

Clearly, healthcare costs have grown since the Center for Medicare and Medicaid Services began tracking this data. It gives more credence to the concerns that people express over its growth. 9.06% CAGR sustained over 55 years is very high.

Source: Proportions paid by each group over time

To answer the first question on sources, "Which payors have faced the most significant in per capita spending growth," I will look at the breakdown of proportions of per-capita healthcare costs paid for by each major payor category.

In [17]:
```
#Copy over the dataframe to make changes without changing the original data
pay_maj_cat_percap = pay_maj_cat.copy()
```

Before looking at the individual categories, we need to convert the numbers to per-capita percentages of total NHE using the formula: $(category / population) / (total / population) * 100$

In [18]:
```
#Need to make #s per capita & percents of the total NHE to see what the non-population related changes are
pay_maj_cat_percap.OUT_OF_POCKET = (pay_maj_cat_percap.OUT_OF_POCKET/pay_maj_cat_percap.POPULATION)/(pay_maj_cat_percap.TOTAL/pay
_maj_cat_percap.POPULATION)*100
pay_maj_cat_percap.INSURANCE = (pay_maj_cat_percap.INSURANCE/pay_maj_cat_percap.POPULATION)/(pay_maj_cat_percap.TOTAL/pay_maj_cat
_percap.POPULATION)*100
pay_maj_cat_percap.OTHER_THIRD_PARTIES = (pay_maj_cat_percap.OTHER_THIRD_PARTIES/pay_maj_cat_percap.POPULATION)/(pay_maj_cat_perc
ap.TOTAL/pay_maj_cat_percap.POPULATION)*100
pay_maj_cat_percap.PUBLIC_HEALTH = (pay_maj_cat_percap.PUBLIC_HEALTH/pay_maj_cat_percap.POPULATION)/(pay_maj_cat_percap.TOTAL/pay
_maj_cat_percap.POPULATION)*100
pay_maj_cat_percap.INVESTMENT = (pay_maj_cat_percap.INVESTMENT/pay_maj_cat_percap.POPULATION)/(pay_maj_cat_percap.TOTAL/pay_maj_c
at_percap.POPULATION)*100
```

Here are the categories' definitions, according to CMS: "

- **Out of pocket**: Includes direct spending by consumers for all health care goods and services, including coinsurance, deductibles, and any amounts not covered by insurance. Premiums paid by individuals for private health insurance are not covered here, but are counted as part of Private Health Insurance.
- **Health insurance**: This aggregated category includes; private health insurance, Medicare, Medicaid, CHIP, Department of Defense, and Department of Veterans Affairs. These plans provide enrollees and beneficiaries insurance against medical losses and, in some instances, directly provide medical care.
- **Other third party payors and programs**: This aggregated category include: worksite health care, other private revenues (e.g. donations), Indian Health Services, workers' compensation, general assistance, maternal/Child Health, vocational rehabilitation, other federal/state and local Programs, SAMHSA (i.e. mental health and substance abuse), and school health
- **Public Health Activity**: In addition to funding the care of individual citizens, the US government is involved in organizing and delivering publicly provided health services such as epidemiological surveillance, inoculations, immunization/vaccination services, disease prevention programs, the operation of public health laboratories, and other such functions.
- **Investement**: This category includes noncommercial research, structures (e.g. new construction of clinics), and equipment "

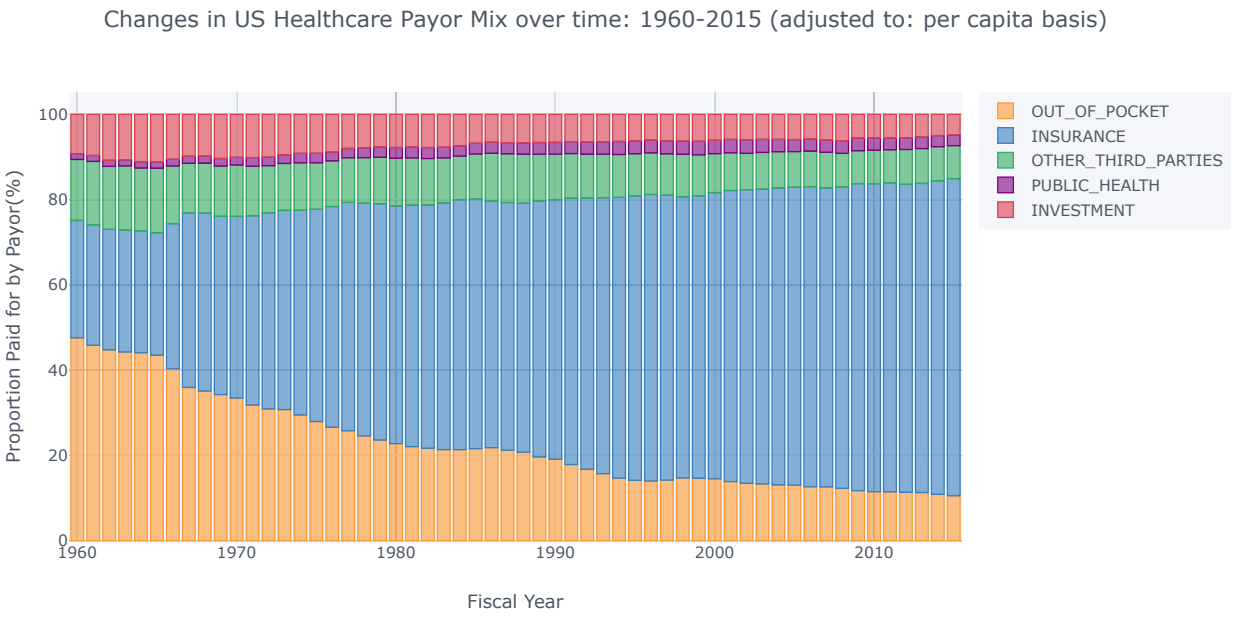
Source: CMS, Quick Definitions for National Health Expenditure Accounts (NHEA) Categories (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/quickref.pdf)

I dropped the total and population variables because they're not relevant right now to graph.

In [19]:
```
pay_maj_cat_percap = pay_maj_cat_percap.drop(['TOTAL','POPULATION'], axis = 1)
```

The stacked bar chart below shows how differrent payors' contributions to healthcare spending have changed over time, adjusting for the changes to population size from 1960 to 2015.

In [20]:
```
#Create stacked bar charts for major categories
pay_maj_cat_percap.iplot(
                        title= 'Changes in US Healthcare Payor Mix over time: 1960-2015 (adjusted to: per capita basis)',
                        xTitle = 'Fiscal Year',
                        yTitle = 'Proportion Paid for by Payor(%)',
                        kind='bar',
                        barmode='stack'
                        )
```



The graph shows **the substitution of out-of-pocket spending by various forms of health insurance.** Specifically, out-of-pocket expenditures have dropped from 48% to 11%. Expenditures paid by other third party payors and programs (mostly by local governments and nonprofits) have dropped from 14% to 8%. Meanwhile, expenditures paid by public and private insurers have nearly tripled from 28% to 74%. **Health insurance** expenditures have grown the most from 1960 to 2015, and this group pays for the largest portion of health care expenditures in the US. These health insurers would be the most interested in engaging with cost-cutting programs and initiatives. Thus, we look into health insurance category to understand what's going on in this important payor group over time.

Exhibit 6. Crosswalk of National Health Expenditure Payers to Business, Household, and Government Sponsors

Sponsor / Payers	Business, Household, and Other Private			Government	
	Private Business	Household	Other Private	Federal	State and Local
Out-of-pocket		X			
Private Health Insurance	X	X		X	X
Other Private Revenues[1]			X		
Medicare	X	X		X	X
Medicaid				X	X
Other Payers[2]	X			X	X

[1] Includes Philanthropy, Private Research, Private Structures and Equipment, and Other Non-Patient revenues.

[2] Includes DOD, VA, CHIP, Worksite Health Care, IHS, Workers' Compensation, general assistance, MCH, vocational rehabilitation, SAMHSA, school health, public health activities, federal and state and local research, and structures and equipment and other federal and state and local programs.

The chart from CMS's methodology paper (above) shows more in more detail who is ultimately footing the bill for each payor group. While there has apparently been a substitution of out-of-pocket spending by private health insurance, it is unclear how that trickles down to the real "sponsor." Therefore, there is even more reason to look deeper into this insurance payment data.

Additional investigation into "health insurance"

Since health insurers pay for such a large portion of national health expenditures, it is worthwhile to examine changes within this category.

Again, I dropped the total and population variables, since they're not of interest at the moment.

```
In [21]: insurance_maj_cat = insurance.drop(['TOTAL_INSURANCE','POPULATION'], axis = 1)
```

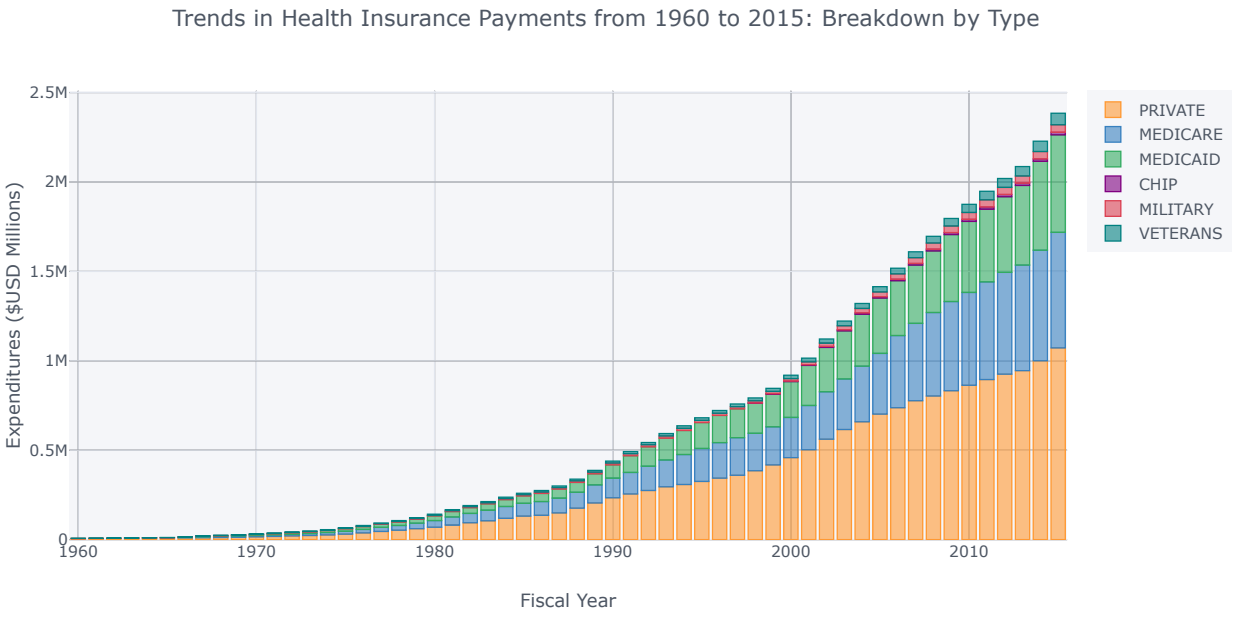
Below is a plot of health insurance payments by the type of insurer.

Here are some definitions of different health insurance payors, according to CMS:"

- Private: UnitedHealth, Aetna and other health insurers who offer plans for individuals and groups (companies, schools, etc.)
- Medicare: Federal public insurance for people aged 65+ (distributed through various private insurers)
- Medicaid: Joint federal and state public insurance for low-income and disabled people
- CHIP: Joint federal and state children's health insurance
- Military: Department of Defense insurance for its employees (e.g. soldiers, etc.)
- Veterans: Department of Veterans Affairs insurance for veterans who served in the armed forces "

Source: CMS, Quick Definitions for National Health Expenditure Accounts (NHEA) Categories (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/quickref.pdf)

In [22]:
```
insurance_maj_cat.iplot(
                        title='Trends in Health Insurance Payments from 1960 to 2015: Breakdown by Type',
                        xTitle = 'Fiscal Year',
                        yTitle = 'Expenditures ($USD Millions)',
                        kind='bar',
                        barmode='stack'
                        )
```



Clearly, **private health insurance remains the biggest insurer group** and payments by this payor has grown over time, and each insurer group has spent more.

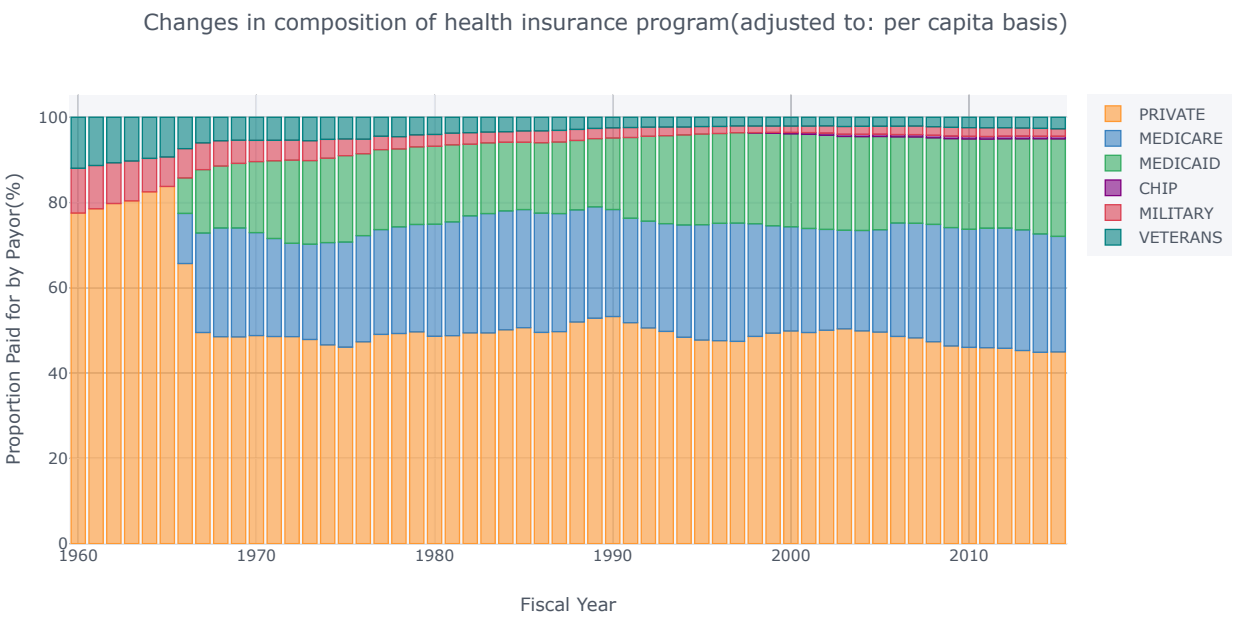
Before looking at the insurer type, we need to convert the numbers to per-capita percentages of total insurance spending using the formula: $(insurer.type / population) / (total.insurance.spending/ population) * 100$

In [23]:
```
#Create a copy to avoid editing the underlying, original dataframe
insurance_percap = insurance.copy()
#Need to make #s per capita to see what the non-population related changes are
insurance_percap.PRIVATE = (insurance_percap.PRIVATE/insurance_percap.POPULATION)/(insurance_percap.TOTAL_INSURANCE/insurance_per
cap.POPULATION)*100
insurance_percap.MEDICARE = (insurance_percap.MEDICARE/insurance_percap.POPULATION)/(insurance_percap.TOTAL_INSURANCE/insurance_pe
rcap.POPULATION)*100
insurance_percap.MEDICAID = (insurance_percap.MEDICAID/insurance_percap.POPULATION)/(insurance_percap.TOTAL_INSURANCE/insurance_pe
rcap.POPULATION)*100
insurance_percap.CHIP = (insurance_percap.CHIP/insurance_percap.POPULATION)/(insurance_percap.TOTAL_INSURANCE/insurance_percap.PO
PULATION)*100
insurance_percap.MILITARY = (insurance_percap.MILITARY/insurance_percap.POPULATION)/(insurance_percap.TOTAL_INSURANCE/insurance_pe
rcap.POPULATION)*100
insurance_percap.VETERANS = (insurance_percap.VETERANS/insurance_percap.POPULATION)/(insurance_percap.TOTAL_INSURANCE/insurance_pe
rcap.POPULATION)*100
```

I dropped the total and population variables because they're not relevant right now to graph.

In [24]:
```
insurance_percap = insurance_percap.drop(['TOTAL_INSURANCE','POPULATION'], axis = 1)
```

In [25]:
```
insurance_percap.iplot(
                        title= 'Changes in composition of health insurance program(adjusted to: per capita basis)',
                        xTitle = 'Fiscal Year',
                        yTitle = 'Proportion Paid for by Payor(%)',
                        kind='bar',
                        barmode='stack'
                        )
```



The legislations behind Medicare and Medicaid were passed in 1965 as part of President Lyndon B. Johnson's "Great Society" program, which seem to have replaced a large portion of private insurance. Dept. of Veterans Affairs and Dept. of Defense expenses have decreased since 1960, which is probably because the largest waves of veterans (WWII, Korean War, Vietnam war) have slowly died off and there have not been similarly large drafting of soldiers ever since. Looking at the figure above, one might draw the conclusion that Medicaid and Medicare **replaced** private health insurance spending. The same person might also take this data and conclude that federal public insurance programs (Medicaid and Medicare) are unnecessary. Therefore, we should actually try to understand what's going on during the first decade of this data (1960 to 1970 because we can then see if there was indeed evidence of strict substitution.

Additional investigation into the per-capita spending from 1960 to 1970

I copy over the data and select for the 1960 to 1970 data that is interesting. Total insurance is not needed to get at the potential substitution.

In [26]:
```
insurance2 = insurance.copy() # make a copy to not change the underly dataset
insurance2 = insurance2.loc['1960':'1970'] #call for only the time period we're interested in at the moment
```

Before looking at the individual categories, we need to convert the numbers to per-capita numbers the following formula: $category / population$

In [27]:
```
#Make payments per capita
insurance2.PRIVATE = insurance2.PRIVATE/insurance2.POPULATION
insurance2.MEDICARE = insurance2.MEDICARE/insurance2.POPULATION
insurance2.MEDICAID = insurance2.MEDICAID/insurance2.POPULATION
insurance2.CHIP = insurance2.CHIP/insurance2.POPULATION
insurance2.MILITARY = insurance2.MILITARY/insurance2.POPULATION
insurance2.VETERANS = insurance2.VETERANS/insurance2.POPULATION
insurance2.TOTAL_INSURANCE = insurance2.TOTAL_INSURANCE/insurance2.POPULATION
```

Before seeing the spending by insurer group, it's useful to look at the growth rates before and after, so I calculate CAGR of per-capital private and total insurance payments growth.

In [28]:
```
PRE_CAGR_PRIVATE = (float(insurance2['PRIVATE'].iloc[5])/float(insurance2['PRIVATE'].iloc[0]))**(1/(1965-1960))-1
PRE_CAGR_TOTAL = (float(insurance2['TOTAL_INSURANCE'].iloc[5])/float(insurance2['TOTAL_INSURANCE'].iloc[0]))**(1/(1965-1960))-1
POST_CAGR_PRIVATE = (float(insurance2['PRIVATE'].iloc[10])/float(insurance2['PRIVATE'].iloc[5]))**(1/(1965-1960))-1
POST_CAGR_TOTAL = (float(insurance2['TOTAL_INSURANCE'].iloc[10])/float(insurance2['TOTAL_INSURANCE'].iloc[5]))**(1/(1965-1960))-1
```

In [29]:
```
print('Growth rate of private insurance, 1960-1965 = ',round(PRE_CAGR_PRIVATE*100,2),'%',';',
      'Growth rate of total insurance, 1960-1965 = ',round(PRE_CAGR_TOTAL*100,2),'%','\n',
      'Growth rate of private insurance, 1965-1970 = ',round(POST_CAGR_PRIVATE*100,2),'%',';',
      'Growth rate of total insurance, 1965-1970 = ',round(POST_CAGR_TOTAL*100,2),'%''\n',
     )
```

```
Growth rate of private insurance, 1960-1965 =  10.02 % ; Growth rate of total insurance, 1960-1965 =  8.32 %
 Growth rate of private insurance, 1965-1970 =  7.94 % ; Growth rate of total insurance, 1965-1970 =  20.27 %
```

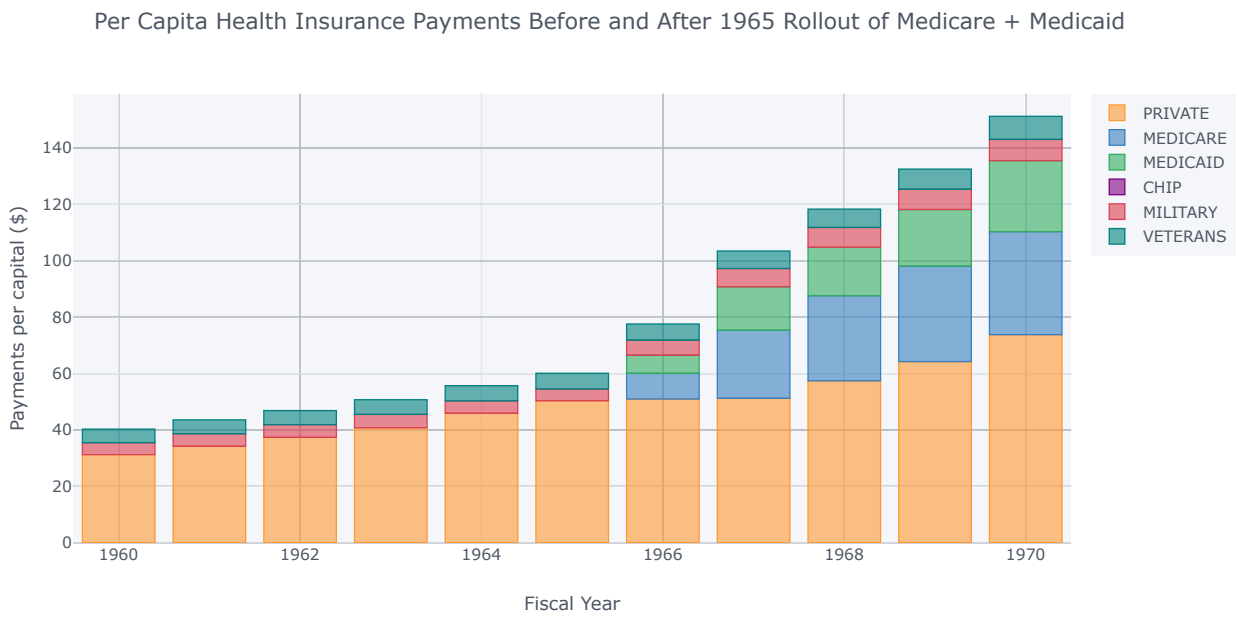
Private insurance payments growth fell from 10% to 8% but total insurance payments growth rate more than doubled from 8% to 20%, perhaps due to the introduction of federal public insurance programs (Medicare and Medicaid).

The population and total insurance variables are not useful to graph, so I dropped them.

In [30]:
```
insurance2 = insurance2.drop(['POPULATION','TOTAL_INSURANCE'], axis = 1)
```

Below is the graph of per-capita health insurance payments by insurer group.

In [31]:
```
insurance2.iplot(
                title='Per Capita Health Insurance Payments Before and After 1965 Rollout of Medicare + Medicaid',
                xTitle = 'Fiscal Year',
                yTitle = 'Payments per capital ($)',
                kind='bar',
                barmode='stack'
                )
```



Interim Conclusion: Similarly, a "zoomed in" look at 1960s per-capita payments by various health insurance payors shows that Medicare & Medicaid did not **replace** the existing services. Rather, it made more healthcare expenditures possible, though it may have taken away some of the growth that health insurance would have had, as evidenced by the changes in growth rates.

Use: Proportions used for each product/service segment over time

Here are the segments' definitions:

- **Hospital Care**: Covers all services provided by hospitals to patients. These include room and board, ancillary charges, services of resident physicians, inpatient pharmacy, hospital-based nursing home and home health care, and any other services billed by hospitals in the United States
- **Physician and Clinical Services**: Covers services provided in establishments operated by Doctors of Medicine (M.D.) and Doctors of Osteopathy (D.O.), outpatient care centers, plus the portion of medical laboratories services that are billed independently by the laboratories
- **Other Professional Services**: Covers services provided in establishments operated by health practitioners other than physicians and dentists
- **Dental Services**: Covers services provided in establishments operated by a Doctor of Dental Medicine (D.M.D.) or Doctor of Dental Surgery (D.D.S.) or a Doctor of Dental Science (D.D.Sc.)
- **Other Health, Residential, and Personal Care**: This category includes spending for Medicaid home and community based waivers, care provided in residential care facilities, ambulance services, school health and worksite health care. Generally these programs provide payments for services in non-traditional settings such as community centers, senior citizens centers, schools, and military field stations
- **Home Health Care**: Covers medical care provided in the home by freestanding home health agencies (HHAs)
- **Nursing Care Facilities and Continuing Care Retirement Communities**: Covers nursing and rehabilitative services provided in freestanding nursing home facilities
- **Prescription Drugs**: Covers the “retail” sales of human-use dosage-form drugs, biological drugs, and diagnostic products that are available only by a prescription
- **Durable Medical Equipment**: Covers “retail” sales of items such as contact lenses, eyeglasses and other ophthalmic products, surgical and orthopedic products, hearing aids, wheelchairs, and medical equipment rentals.
- **Other Non-Durable Medical Products**: Covers the “retail” sales of non-prescription drugs and medical sundries.
- **Total Administration and Total Net Cost of Health** : Covers the cost of administration by the government and insurers
- **Public Health Activity**: In addition to funding the care of individual citizens, the US government is involved in organizing and delivering publicly provided health services such as epidemiological surveillance, inoculations, immunization/vaccination services, disease prevention programs, the operation of public health laboratories, and other such functions.
- **Research**: This category includes noncommercial research.
- **Total Structures and equipment** : This category includes structures (e.g. new construction of clinics) and equipment

Source: CMS, Quick Definitions for National Health Expenditure Accounts (NHEA) Categories (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/quickref.pdf)

I first made a separate copy of the segmeents dataframe and converted the data into percentages of per-capita spending by category using the following formula: $(segment / population) / (total/ population) * 100$

In [32]:
```
#Make a copy
segments_percap = segments.copy()
#Need to make #s per capita to see what the non-population related changes in segment proportions are
segments_percap.HOSPITAL = (segments_percap.HOSPITAL/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATIO
N)*100
segments_percap.PHYSICIAN = (segments_percap.PHYSICIAN/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULAT
ION)*100
segments_percap.DENTAL = (segments_percap.DENTAL/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATION)*1
00
segments_percap.SERVICES = (segments_percap.SERVICES/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATIO
N)*100
segments_percap.HOME_CARE = (segments_percap.HOME_CARE/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULAT
ION)*100
segments_percap.PRODUCTS = (segments_percap.PRODUCTS/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATIO
N)*100
segments_percap.DRUGS = (segments_percap.DRUGS/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATION)*100
segments_percap.EQUIPMENT = (segments_percap.EQUIPMENT/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULAT
ION)*100
segments_percap.NURSING = (segments_percap.NURSING/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATION)
*100
segments_percap.OTHER = (segments_percap.OTHER/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATION)*100
segments_percap.ADMIN = (segments_percap.ADMIN/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATION)*100
segments_percap.PUBLIC_HEALTH = (segments_percap.PUBLIC_HEALTH/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap
.POPULATION)*100
segments_percap.RESEARCH = (segments_percap.RESEARCH/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_percap.POPULATIO
N)*100
segments_percap.INFRASTRUCTURE = (segments_percap.INFRASTRUCTURE/segments_percap.POPULATION)/(segments_percap.TOTAL/segments_perc
ap.POPULATION)*100
```

Since there are way too many categories to keep track of, I group together other health, residential, and personal care, administrative, public health activity, research, and infrastructure investment costs as "non-core."

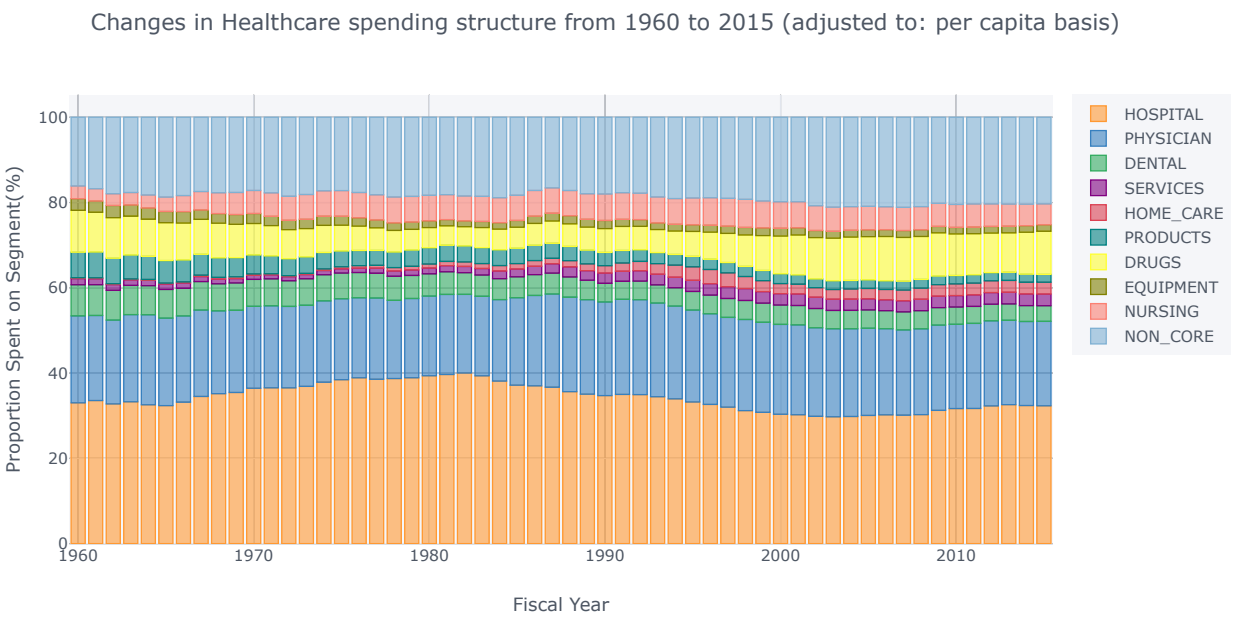
In [33]:
```
segments_percap['NON_CORE']=list(segments_percap.OTHER+segments_percap.ADMIN+segments_percap.PUBLIC_HEALTH+segments_percap.RESEAR
CH+segments_percap.INFRASTRUCTURE)
```

I dropped the total and population variables because they're not relevant right now to graph.

In [34]:
```
segments_percap1 = segments_percap.drop(['TOTAL','POPULATION','OTHER','ADMIN','PUBLIC_HEALTH','RESEARCH','INFRASTRUCTURE']
                                        , axis = 1)
```

Below is a graph of historical spending broken down by service/product segments

In [35]:
```
segments_percap1.iplot(
                        title= 'Changes in Healthcare spending structure from 1960 to 2015 (adjusted to: per capita basis)',
                        xTitle = 'Fiscal Year',
                        yTitle = 'Proportion Spent on Segment(%)',
                        kind='bar',
                        barmode='stack'
                        )
```



The two biggest categories of spending are persistently **hospital-related expenditures and physician and clinical services.** In 2015, 32% and 20%, respectively, of total national health expenditures were spent on these to categories. This is unsurprising as we spend the most of our health dollars on these two main providers of healthcare.

Retail prescription drugs follows behind physician-realted spending as the third largest category spent on. However, for all of the attention brought to this category, it is still a relatively small portion of the total spending, though it has nearly doubled. In fact, since 1990, the most significant increase has been in proportion of expenditures from the retail prescription drug sales segment (from 6 to 10%). Meanwhile, the proportion of expenditures from the hospital segment has been decreasing since 1980 (from 39% to 32%).

Non-core expenditures have not seen the type of proportional increase since the roll-out of the Affordable Care Act on October 1, 2013. Doctors have complained that they are facing a lot of "administrative costs" , but the costs of administration of health insurance have not significantly changed. While one might attribute the former observation to the wrong categorization, the costs of physician and clinical services and hospital remain relatively the same. To further disprove the claim, one would need additional years of data.

The hospital and clinical services segments remain the biggest opportunities for cost-cutting. Another segment to closely monitor is the drugs expense, which has grown quickly.

5. Discussion

Source Question: Who has paid for the US's growing national health expenditures

- **Answer:** Since 1960, the biggest trend has been the substitution of out-of-pocket and perhaps other third-party payors' healthcare spending by various health insurers. Private health insurance makes up the biggest group and has grown significantly.
- **Interesting insight:** The introduction of Medicaid and Medicare did not seem to replace the existing chunks of served spending. Rather it seemed to make possible additional spending rather than purely cannibalizing the existing health insurance spending.

Use Question: What product/service segments has the US spent said money on?

- **Answer:** More than half of expenditures have been consistently spent at the hospital or physician-led clinics. This is not suprising.
- **Intersting insight:** Prescription drugs have increasingly made up a larger portion of total healthcare spending. However, it does not make up as large a portion as conversations have made it out to seem.

Next Steps for Investigators: There are many related investigations that I did not have time to get to that would be interesting to look at.

- **Add on to this data set with historical outcomes data:** Get at the "bang for buck." In other words, are we paying more for the same, lower, or higher level of health? One quantitative measure to look at could be average lifespan or any of the other health indicators made for health rankings.
- **State Health Expenditure Accounts (SHEA) data (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsStateHealthAccountsProvider.html)**: This dataset presents "aggregate and per capita estimates of health care spending by type of establishment delivering care (hospitals, physicians and clinics, nursing homes, etc.) and for medical products (prescription drugs, over-the-counter medicines and sundries and durable medical products such as eyeglasses and hearing aids) purchased in retail outlets." With states-level data, one can get at differences in spending and payor mix. If one has additional data on different healthcare systems within those states, one can then run regressions that evaluate what drives those differences in spending and spending growth, further fleshing out the uses section.
- **Personal health care (PHC) spending (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Age-and-Gender.html)** This information is most relevant to patients/voters. CMS provides costs data broken down by type of good or service and is available bi-annually for five age groups: 0-18, 19-44, 45-64, 65-84, and 85 and over and for males and females from 2002 through 2012. This can get at the question: How do growth drivers (segments with the highest growth) differ across age group and gender dimensions? Do gender or age really matter in terms of cost levels?