

R Markdown

INFO 201

Today's Objectives

Practice working with a REST API

Consider the importance of generating **dynamic reports**

Learn how to build dynamic reports with **R Markdown**

Introduce the concept of **GitHub Branches**

API Review

REST APIs

Representational State Transfer APIs

Exposes **data** components

Transfers information with HTTP (HyperText Transfer Protocol)

Enables querying content over the web

The New York Times Developer Network

All the APIs Fit to POST

You already know that NYTimes.com is an unparalleled source of news and information. But now it's a premier source of data, too — why just read the news when you can hack it?

Getting Started

The Times Developer Network is our API clearinghouse and community. Here's how to get started:

1. Request an [API key](#)
2. Read the API documentation, [FAQ](#) and [Terms of Use](#)
3. Use the API Tool associated with each API to experiment without writing code

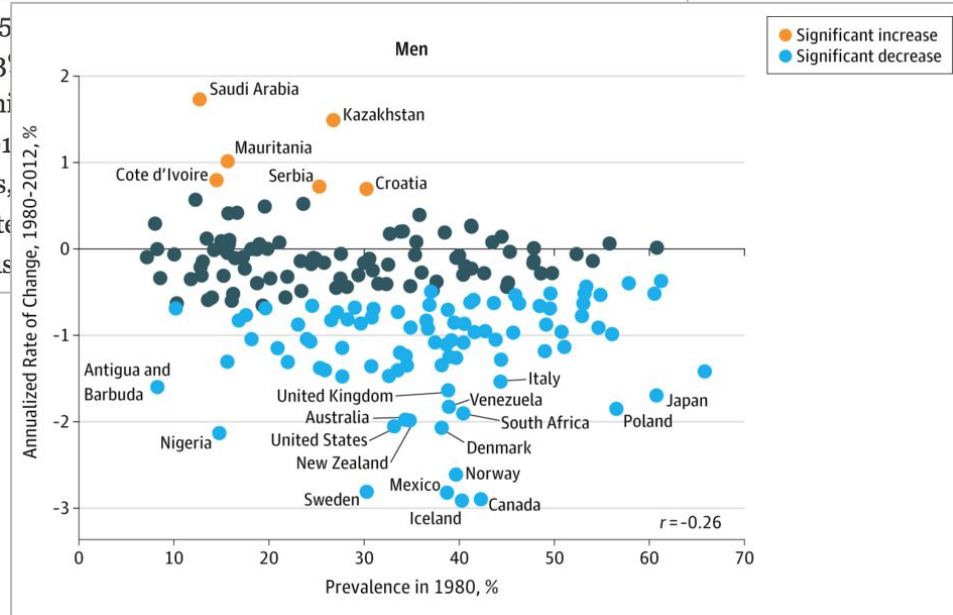
module 10 exercise-5

Dynamic Reports

Results Global modeled age-standardized prevalence of daily tobacco smoking in the population older than 15 years decreased from 41.2% (95% uncertainty interval [UI], 40.0%-42.6%) in 1980 to 31.1% (95% UI, 30.2%-32.0%; $P < .001$) in 2012 for men and from 10.6% (95% UI, 10.2%-11.1%) to 6.2% (95% UI, 6.0%-6.4%; $P < .001$) for women. Global modeled prevalence declined at a faster rate from 1996 to 2006 (mean annualized rate of decline, 1.7%; 95% UI, 1.5%-1.9%) compared with the subsequent period (mean annualized rate of decline, 0.9%; 95% UI, 0.5%-1.3%; $P = .003$). Despite the decline in modeled prevalence, the number of daily smokers increased from 721 million (95% UI, 700 million–742 million) in 1980 to 967 million (95% UI, 944 million–989 million; $P < .001$) in 2012. Modeled prevalence rates exhibited substantial variation across age, sex, and countries, with rates below 5% for women in some African countries to more than 55% for men in Timor-Leste and Indonesia. The number of cigarettes per smoker per day also varied widely across countries and was not correlated with modeled prevalence.

How many numbers are in this summary?

Results Global modeled age-standardized prevalence of daily tobacco smoking in the population older than 15 years decreased from 41.2% (95% uncertainty interval [UI], 40.0%-42.6%) in 1980 to 31.1% (95% UI, 30.2%-32.0%; $P < .001$) in 2012 for men and from 10.6% (95% UI, 10.2%-11.1%) to 6.2% (95% UI, 6.0%-6.4%; $P < .001$) for women. Global modeled prevalence declined at a faster rate from 1996 to 2006 (mean annualized rate of decline, 1.7%; 95% UI, 1.5%-1.9%) than from 1980 to 2006 (mean annualized rate of decline, 0.9%; 95% UI, 0.5%-1.3%). The number of daily smokers increased from 721 million (95% UI, 644 million–989 million; $P < .001$) in 1980 to 1.1 billion (95% UI, 944 million–1.2 billion) in 2012. There was substantial variation across age, sex, and countries, with prevalence ranging from less than 5% for women in Timor-Leste to more than 55% for men in Timor-Leste. The number of cigarettes smoked per day also varied widely across countries and was



Reports have hundreds (thousands) of values, dozens of graphics

How can we manage
updating reports as
our analysis
changes?

```

10 avg_arrival_delay <- flights %>%
11   group_by(dest) %>%
12   summarise(avg_delay = mean(arr_delay))
13 mutate(faa = dest) %>%
14 left_join(airports,
15   arrange(-avg_delay))
16
17 # Create a dataframe of the average arrival delay by destination
18 # to join on the "airlines" data
19 avg_airline_delay <- flights %>%
20   group_by(airline) %>%
21   summarise(avg_delay = mean(arr_delay))
22 left_join(airlines,
23   arrange(-avg_delay))
24
25 ### Bonus ###
26 # Calculate the average arrival delay by city, then merge on the
27 avg_city_airline <- flights %>%
28   group_by(city) %>%
29   summarise(avg_delay = mean(arr_delay, na.rm=TRUE)) %>%

```

20:40 (Top Level) :

Console ~/Documents/INFO-498F/

+ arrange(-avg_delay)

> avg_arrival_delay

Source: local data frame [105 x 5]

	dest	avg_delay	faa	lon	alt	tz	dst
1	CAE	41.76415	CAE	-81.11953	236	-5	A
2	TUL	33.65986	TUL	-85.88811	677	-6	A
3	OKC	30.61905	OKC	-100.073	1295	-6	A
4	JAC	28.09524	JAC	-75.6451	-7	A	
5	TYS	24.06920	TYS	-81.281	-5	A	
6	MSN	20.19604	MSN	-88.625	-5	A	
7	RIC	20.11125	RIC	-77.043	-5	A	
8	CAK	19.69834	CAK	-83.547	-5	A	
9	DSM	19.00574	DSM	-93.625	-5	A	
10	GRR	18.18956	GRR	-85.547	-5	A	

Between 1980 and 2012, the age-standardized prevalence of daily tobacco smoking for men declined from 31.1% (95% UI, 30.2%-32.0%; $P < .001$), an average annual decline of 0.1% (95% UI, 0.0%-0.1%; $P < .001$) and for women declined from 10.6% (95% UI, 10.0%-11.2%; $P < .001$) to 6.4% (95% UI, 6.0%-6.4%; $P < .001$) or 1.7% (95% UI, 1.5%-1.9%; $P < .001$) per year. Progress in reducing the age-standardized prevalence of smokers has been modest for both men and women (Figure 1B): modest progress from 1980 to 1990 (annual rate of decline, 0.6%; 95% UI, 0.4%-0.8%), followed by a decade of modest progress (annual rate of decline, 1.7%; 95% UI, 1.5%-1.9%; $P < .001$), then a period of apparent increase since 2010 for men. This deceleration in the rate of decline is apparent in several large countries, including China, Indonesia, and Russia. Since 1980, the global rate of decline has been higher than in men. eTable 11 in the Supplement provides various standardized prevalence rates, rates of decline, and total cigarettes consumed each year worldwide. While estimates of the number of men and women who smoke daily, increasing from 1.1 billion (95% UI, 944 million) in 1980 to 1.6 billion (95% UI, 1.44 billion) in 2012, the number of cigarettes smoked declined from 6.25 trillion (95% UI, 5.6 trillion) in 1980 to 4.78 trillion (95% UI, 4.78 trillion-5.16 trillion) in 2012. There was no discernible change in the number of cigarettes smoked per smoker per day, remaining around

R Markdown

R Markdown

A way to generate dynamic reports

Perform analysis and build report in the **same document**

Allows you to populate text with values from your analysis

Seamlessly integrate text and graphics

Top Level Header

Second Level Header

```

block of code  
across multiple lines

```

- List item 1
- List item 2
- List item 3

>Here is a block quote.

Top Level Header

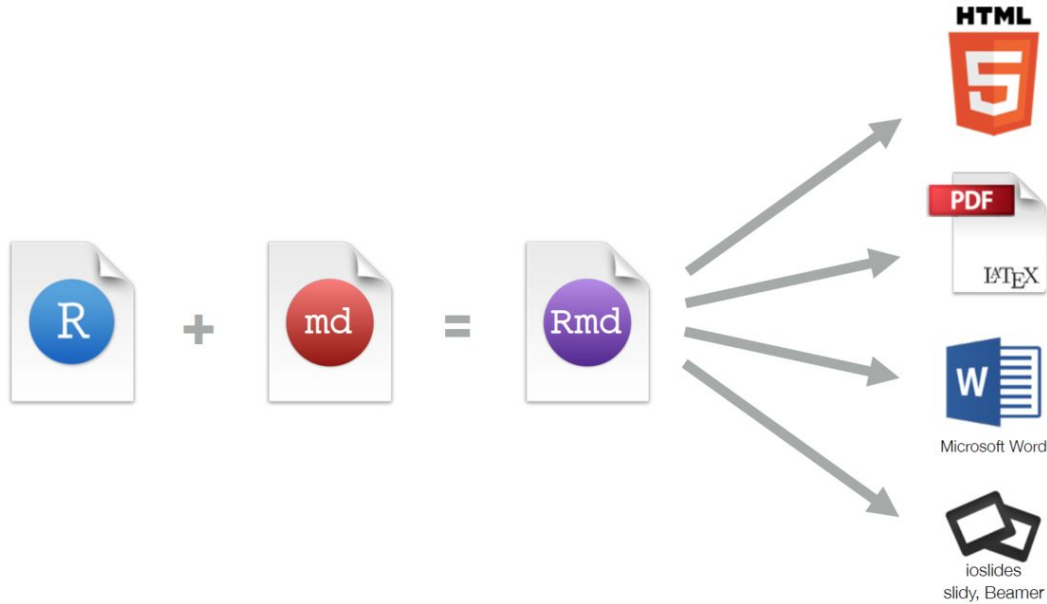
Second Level Header

block of code
across multiple lines

- List item 1
- List item 2
- List item 3

Here is a block quote.

R Markdown

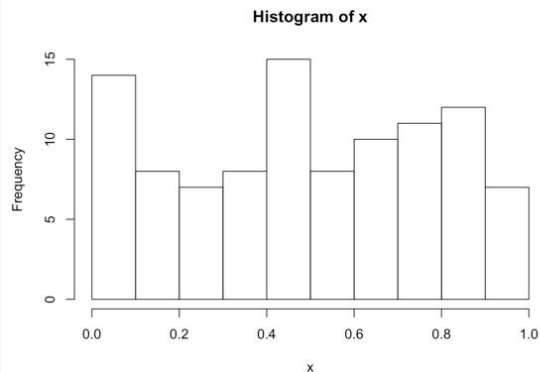


This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

```
```\r\nx <- runif(1:100)\r\nhist(x)\r\n```
```

This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

```
x <- runif(1:100)\r\nhist(x)
```





This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

```
```${r, eval=FALSE}  
x <- runif(1:100)  
hist(x)  
```
```

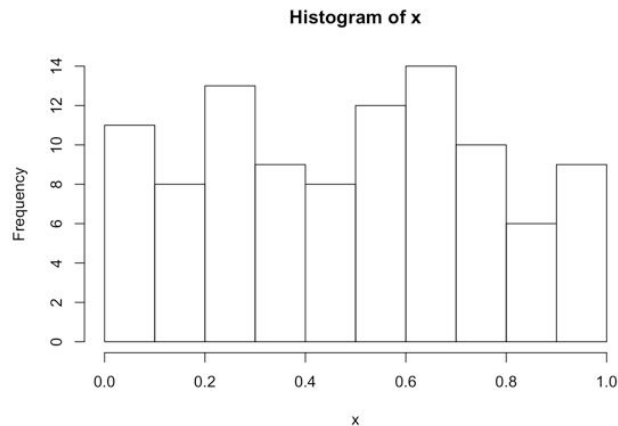
This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

```
x <- runif(1:100)
hist(x)
```

This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

```
```{r, echo=FALSE}
x <- runif(1:100)
hist(x)
```
```

This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

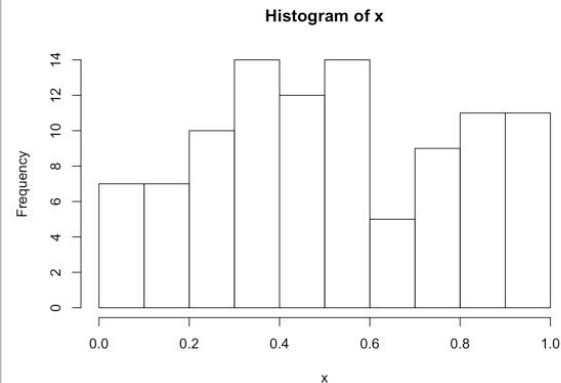


This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!

```
```{r, echo=FALSE}
x <- runif(1:100)
hist(x)
x.mean <- mean(x)
```
```

You can reference a value in-line like this: ``r x.mean``.

This is the code we will look at in class. This is just plain old Markdown that lets you render text in **bold** or *italics*. However, we can put in a chunk or R code, and it will show us the code and results!



You can reference a value in-line like this: 0.5173921.

module 11 exercise-1

# Git Branches

---

A **branch** in Git is a way of labeling a ***sequence of commits***. You can create labels (branches) for different commits, and effectively have ***different "lines" of development*** occurring in parallel and diverging from each other.

# Use of Branches

Allows parallel lines of development

Can focus on new features in a testing environment

Facilitates collaboration

# Additional benefits....

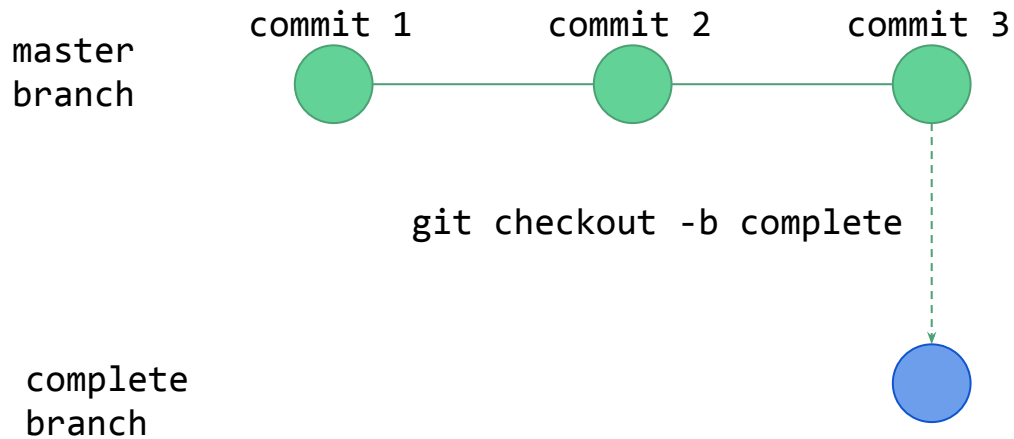
GitHub will host webpages for **gh-pages** branch

If we create a .html file from R Markdown, that is a website!

The file named index.html will be displayed automatically at a web URL

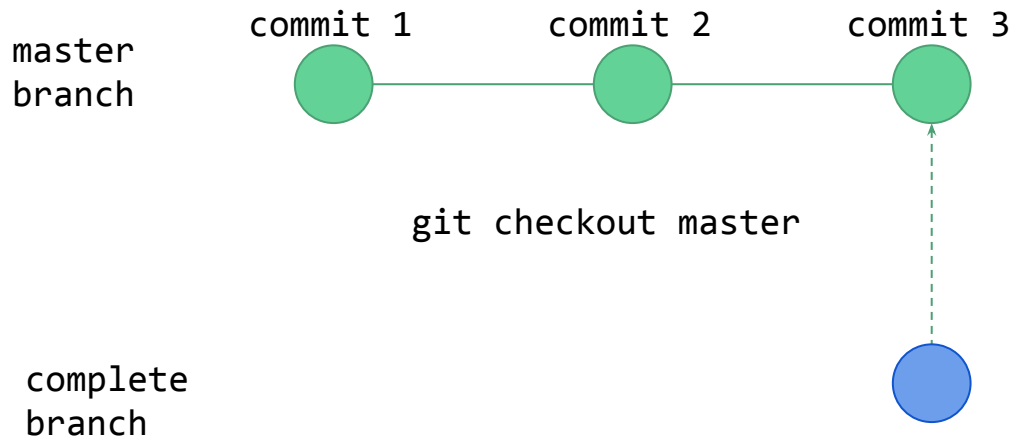
This will be our current focus for now (creating a gh-pages branch, keeping it the same as your master branch)





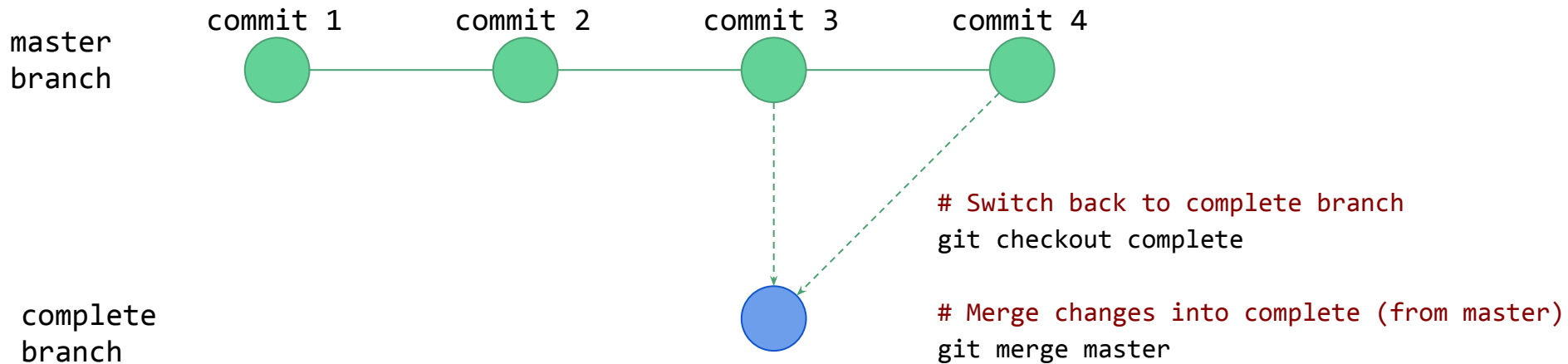
Create a new branch from your current (master) branch

[module-11](#)



Checkout an existing (master) branch (from complete branch)

[module-11](#)



# Create gh-pages branch from master branch

```
git checkout -b gh-pages
```

# Switch back to master branch: I suggest you always develop in your master branch

```
git checkout master
```

# Make some changes, then add and commit as usual

```
git add .
```

```
git commit -m "Made more changes here on the master branch"
```

# Switch back to your gh-pages branch

```
git checkout gh-pages
```

# Merge in your changes into your gh-pages branch from your master branch

```
git merge master
```

# Push all changes up to GitHub

```
git push origin --all
```

# Upcoming...

By Tuesday: Be confident with **module 11**

Due Tuesday, 11/1 (***before class***): [a5-github-report](#)