



Reading JSON

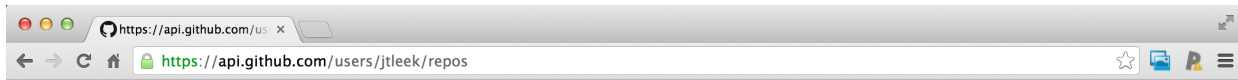
Jeffrey Leek
Johns Hopkins Bloomberg School of Public Health

JSON

- Javascript Object Notation
- Lightweight data storage
- Common format for data from application programming interfaces (APIs)
- Similar structure to XML but different syntax/format
- Data stored as
 - Numbers (double)
 - Strings (double quoted)
 - Boolean (*true* or *false*)
 - Array (ordered, comma separated enclosed in square brackets `[]`)
 - Object (unordered, comma separated collection of key:value pairs in curly brackets `{}`)

<http://en.wikipedia.org/wiki/JSON>

Example JSON file



```
[
  {
    "id": 12441219,
    "name": "ballgown",
    "full_name": "jtleek/ballgown",
    "owner": {
      "login": "jtleek",
      "id": 1571674,
      "avatar_url": "https://avatars.githubusercontent.com/u/1571674?v=4",
      "gravatar_id": "4bd13719da0ba2c5bd2a446e14f78187",
      "url": "https://api.github.com/users/jtleek",
      "html_url": "https://github.com/jtleek",
      "followers_url": "https://api.github.com/users/jtleek/followers",
      "following_url": "https://api.github.com/users/jtleek/following{/other_user}",
      "gists_url": "https://api.github.com/users/jtleek/gists{/gist_id}",
      "starred_url": "https://api.github.com/users/jtleek/starred{/owner}/{repo}",
      "subscriptions_url": "https://api.github.com/users/jtleek/subscriptions",
      "organizations_url": "https://api.github.com/users/jtleek/orgs",
      "repos_url": "https://api.github.com/users/jtleek/repos",
      "events_url": "https://api.github.com/users/jtleek/events{/privacy}",
      "received_events_url": "https://api.github.com/users/jtleek/received_events",
      "type": "User",
      "site_admin": false
    },
    "private": false,
    "html_url": "https://github.com/jtleek/ballgown",
    "description": "Code for manipulating ballgown output in R",
    "fork": true,
    "url": "https://api.github.com/repos/jtleek/ballgown",
    "forks_url": "https://api.github.com/repos/jtleek/ballgown/forks",
    "keys_url": "https://api.github.com/repos/jtleek/ballgown/keys{/key_id}",
    "collaborators_url": "https://api.github.com/repos/jtleek/ballgown/collaborators{/collaborator}",
    "teams_url": "https://api.github.com/repos/jtleek/ballgown/teams",
    "hooks_url": "https://api.github.com/repos/jtleek/ballgown/hooks",
    "issue_events_url": "https://api.github.com/repos/jtleek/ballgown/issues/events{/number}",
    "labels_url": "https://api.github.com/repos/jtleek/ballgown/labels{/name}",
    "merges_url": "https://api.github.com/repos/jtleek/ballgown/merges",
    "milestones_url": "https://api.github.com/repos/jtleek/ballgown/milestones{/number}",
    "notifications_url": "https://api.github.com/repos/jtleek/ballgown/notifications{/thread_id}",
    "pulls_url": "https://api.github.com/repos/jtleek/ballgown/pulls{/number}",
    "releases_url": "https://api.github.com/repos/jtleek/ballgown/releases{/id}",
    "squid_url": "https://api.github.com/repos/jtleek/ballgown/squid{/id}",
    "tags_url": "https://api.github.com/repos/jtleek/ballgown/tags{/name}",
    "trees_url": "https://api.github.com/repos/jtleek/ballgown/trees{/sha}"
  }
]
```

Reading data from JSON {jsonlite package}

```
library(jsonlite)
jsonData <- fromJSON("https://api.github.com/users/jtleek/repos")
names(jsonData)
```

```
[1] "id"          "name"          "full_name"     "owner"
[5] "private"     "html_url"      "description"    "fork"
[9] "url"         "forks_url"     "keys_url"      "collaborators_url"
[13] "teams_url"   "hooks_url"     "issue_events_url" "events_url"
[17] "assignees_url" "branches_url"  "tags_url"      "blobs_url"
[21] "git_tags_url" "git_refs_url"  "trees_url"     "statuses_url"
[25] "languages_url" "stargazers_url" "contributors_url" "subscribers_url"
[29] "subscription_url" "commits_url"   "git_commits_url" "comments_url"
[33] "issue_comment_url" "contents_url"  "compare_url"    "merges_url"
[37] "archive_url"   "downloads_url" "issues_url"     "pulls_url"
[41] "milestones_url" "notifications_url" "labels_url"     "releases_url"
[45] "created_at"    "updated_at"    "pushed_at"     "git_url"
[49] "ssh_url"       "clone_url"     "svn_url"        "homepage"
[53] "size"          "stargazers_count" "watchers_count" "language"
```

Nested objects in JSON

```
names( jsonData$owner)
```

```
[1] "login"           "id"               "avatar_url"       "gravatar_id"
[5] "url"             "html_url"         "followers_url"     "following_url"
[9] "gists_url"       "starred_url"      "subscriptions_url" "organizations_url"
[13] "repos_url"       "events_url"       "received_events_url" "type"
[17] "site_admin"
```

```
jsonData$owner$login
```

```
[1] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
[11] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
[21] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
```

Writing data frames to JSON

```
myjson <- toJSON(iris, pretty=TRUE)
cat(myjson)
```

```
[
  {
    "Sepal.Length" : 5.1,
    "Sepal.Width" : 3.5,
    "Petal.Length" : 1.4,
    "Petal.Width" : 0.2,
    "Species" : "setosa"
  },
  {
    "Sepal.Length" : 4.9,
    "Sepal.Width" : 3,
    "Petal.Length" : 1.4,
    "Petal.Width" : 0.2,
    "Species" : "setosa"
  },
  {
    "Sepal.Length" : 4.7,
```

Convert back to JSON

```
iris2 <- fromJSON(myjson)
head(iris2)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

<http://www.r-bloggers.com/new-package-jsonlite-a-smarter-json-encoderdecoder/>

Further resources

- <http://www.json.org/>
- A good tutorial on jsonlite - <http://www.r-bloggers.com/new-package-jsonlite-a-smarter-json-encoderdecoder/>
- [jsonlite vignette](#)