

flickr predictive analysis



Where do people like to take photographs now?

Where will people take photographs in the future?



Where will people want to go in the future?

- Photo/video identifier
- User NSID, User nickname
- Date taken
- Date uploaded
- Capture device
- Title, Description
- User tags (comma-separated)
- Machine tags (comma-separated)
- Longitude
- Latitude
- Accuracy
- Photo/video page URL, Photo/video download URL
- License name, License URL
- Photo/video server identifier, Photo/video farm identifier
- Photo/video secret, Photo/video secret original
- Photo/video extension original
- Photos/video marker (0 = photo, 1 = video)

```

10201275523 55289779@N00 booledozer 2013-10-09 15:30:13.0 1381464628 Panasonic+DMC-ZS10
          Images+from+the+intersection+of+Parliament+and+Queen%2C+2013+10+09+%283%29
          -79.365637 43.649708 16 http://www.flickr.com/photos/55289779@N00/10201275523/
http://farm6.staticflickr.com/5465/10201275523_3e6ea67c7f.jpg Attribution-ShareAlike License http://
creativecommons.org/licenses/by-sa/2.0/ 5465 6 3e6ea67c7f f400826feb jpg 0
2297552664 8134076@N03 nathalie.cone 2008-02-27 08:14:36.0 1204166915 Canon+PowerShot+S2+IS
denise%27s+peanut+chicken recipe+here%3A+%3Ca+href%3D%22http%3A%2F%2Fallrecipes.com%2FRecipe
%2FDenises-Peanut-Chicken%2FDetail.aspx%22%3Eallrecipes.com%2FRecipe%2FDenises-Peanut-Chicken%2FDetail.aspx
%3C%2Fa%3E%0Ai+added+1%2F2+teaspoon+of+sriracha+hot+chili+sauce+and+used+1+TBS+of+chunky+peanut+butter+in
+place+of+the+2+cups+of+peanuts. http://www.flickr.com/photos/
8134076@N03/2297552664/ http://farm4.staticflickr.com/3171/2297552664_1ee0e8855d.jpg Attribution-
NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/ 3171 4
1ee0e8855d bed1547a29 jpg 0
7289030198 53430201@N03 sarabeephoto 2012-05-28 00:23:30.0 1338233199 Canon+EOS+DIGITAL
+REBEL+XSi blip+fest+2012+night+3
blip,blip12ny,blipfest,blipfestival,blipfestival2012,blipfestivalnewyork,blipfestivalnewyork2012,chipmusic,ch
iptune,chip tunes -73.985495 40.740067 15 http://www.flickr.com/photos/
53430201@N03/7289030198/ http://farm8.staticflickr.com/7231/7289030198_1f1ba44113.jpg Attribution-
NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/ 7231 8
1f1ba44113 bc36823aec jpg 0
4140939180 12940050@N02 enjoiskate8 2009-11-21 08:03:30.0 1259417601 PENTAX+Corporation
+PENTAX+K10D IMGP7409 2009,amsterdam,clinquest,fall,netherlands,nov,terschelling,travel
http://www.flickr.com/photos/12940050@N02/4140939180/ http://
farm3.staticflickr.com/2565/4140939180_07aeded917.jpg Attribution-NonCommercial-NoDerivs License
http://creativecommons.org/licenses/by-nc-nd/2.0/ 2565 3 07aeded917 da7f4b048e jpg
0
9506922316 80547277@N00 eliduke 2012-08-28 07:29:42.0 1376439487 Canon+EOS+Kiss+X3
Christchurch%3A+On+our+way+to+Antarctica%21 ben,cdc,christchurch,new-zealand
http://www.flickr.com/photos/80547277@N00/9506922316/ http://
farm6.staticflickr.com/5468/9506922316_c19019e38f.jpg Attribution-ShareAlike License http://
creativecommons.org/licenses/by-sa/2.0/ 5468 6 c19019e38f 82de2f240e jpg 0
4436463882 42132616@N04 Miriam+Jones 2010-03-15 13:49:02.0 1268686142 San+Antonio
+House+Concert%2C+March+2010 http://www.flickr.com/photos/
42132616@N04/4436463882/ http://farm5.staticflickr.com/4007/4436463882_b96a3d9df9.jpg Attribution-
NoDerivs License http://creativecommons.org/licenses/by-nd/2.0/ 4007 5 b96a3d9df9
d901bc18a1 jpg 0
4572998878 78969707@N00 jkgreenstein12 2010-05-01 16:28:20.0 1272843440 Panasonic+DMC-LS70
Matt+and+Diana+Wedding+weekend Greeny%27s%2C+Christines%2C+Andy%2C+Jeeane+and+Lou%27s+photos+from
+the+wedding+weekend. diana,matt,wedding -71.047843 42.33719 16 http://
www.flickr.com/photos/78969707@N00/4572998878/ http://farm4.staticflickr.com/3373/4572998878_658b45226f.jpg
Attribution-NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/
3373 4 658b45226f a0777fdc0f jpg 0
9329902958 97468058@N00 angela+louise 2013-07-20 19:27:24.0 1374344844 Too+hot+%21
bantockpark,fayne,sean,strawberry,summer2013 http://
www.flickr.com/photos/97468058@N00/9329902958/ http://farm4.staticflickr.com/3819/9329902958_0bc80ce58a.jpg
Attribution-NonCommercial-ShareAlike License http://creativecommons.org/licenses/by-nc-sa/2.0/
3819 4 0bc80ce58a cd1d8e31ed jpg 0
3174965401 36813788@N00 Arthur+2+Sheds 2008-12-31 21:29:32.0 1231293771 Canon+EOS+DIGITAL
+REBEL+XTi IMG_3398 2009,new,years http://
www.flickr.com/photos/36813788@N00/3174965401/ http://farm4.staticflickr.com/3099/3174965401_2742004421.jpg
Attribution-NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/
3099 4 2742004421 486c7d19b5 jpg 0

```

Acquiring Data: YFCC100M Dataset

100 Million Photos + Videos

- Photo/video identifier
- User NSID, User nickname
- **Date taken**
- Date uploaded
- **Capture device**
- Title, Description
- User tags (comma-separated)
- Machine tags (comma-separated)
- **Longitude**
- **Latitude**
- Accuracy
- Photo/video page URL, Photo/video download URL
- License name, License URL
- Photo/video server identifier, Photo/video farm identifier
- Photo/video secret, Photo/video secret original
- Photo/video extension original
- Photos/video marker (0 = photo, 1 = video)

```

10201275523 55289779@N00 booledozer 2013-10-09 15:30:13.0 1381464628 Panasonic+DMC-ZS10
          Images+from+the+intersection+of+Parliament+and+Queen%2C+2013+10+09+%283%29
          -79.365637 43.649708 16 http://www.flickr.com/photos/55289779@N00/10201275523/
http://farm6.staticflickr.com/5465/10201275523_3e6ea67c7f.jpg Attribution-ShareAlike License http://
creativecommons.org/licenses/by-sa/2.0/ 5465 6 3e6ea67c7f f400826feb jpg 0
2297552664 8134076@N03 nathalie.cone 2008-02-27 08:14:36.0 1204166915 Canon+PowerShot+S2+IS
denise%27s+peanut+chicken recipe+here%3A+%3Ca+href%3D%22http%3A%2F%2Fallrecipes.com%2FRecipe
%2FDenises-Peanut-Chicken%2FDetail.aspx%22%3Aallrecipes.com%2FRecipe%2FDenises-Peanut-Chicken%2FDetail.aspx
%3C%2Fa%3E%0Ai+added+1%2F2+teaspoon+of+sriracha+hot+chili+sauce+and+used+1+TBS+of+chunky+peanut+butter+in
+place+of+the+2+cups+of+peanuts. http://www.flickr.com/photos/
8134076@N03/2297552664/ http://farm4.staticflickr.com/3171/2297552664_1ee0e8855d.jpg Attribution-
NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/ 3171 4
1ee0e8855d bed1547a29 jpg 0
7289030198 53430201@N03 sarabeephoto 2012-05-28 00:23:30.0 1338233199 Canon+EOS+DIGITAL
+REBEL+XSi blip+fest+2012+night+3
blip,blip12ny,blipfest,blipfestival,blipfestival2012,blipfestivalnewyork,blipfestivalnewyork2012,chipmusic,ch
iptune,chip tunes -73.985495 40.740067 15 http://www.flickr.com/photos/
53430201@N03/7289030198/ http://farm8.staticflickr.com/7231/7289030198_1f1ba44113.jpg Attribution-
NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/ 7231 8
1f1ba44113 bc36823a2c jpg 0
4140939180 12940050@N02 enjoiskate8 2009-11-21 08:03:30.0 1259417601 PENTAX+Corporation
+PENTAX+K10D IMGP7409 2009,amsterdam,clinquest,fall,netherlands,nov,terschelling,travel
http://www.flickr.com/photos/12940050@N02/4140939180/ http://
farm3.staticflickr.com/2565/4140939180_07aeded917.jpg Attribution-NonCommercial-NoDerivs License
http://creativecommons.org/licenses/by-nc-nd/2.0/ 2565 3 07aeded917 da7f4b048e jpg
0
9506922316 80547277@N00 eliduke 2012-08-28 07:29:42.0 1376439487 Canon+EOS+Kiss+X3
Christchurch%3A+On+our+way+to+Antarctica%21 ben,cdc,christchurch,new-zealand
http://www.flickr.com/photos/80547277@N00/9506922316/ http://
farm6.staticflickr.com/5468/9506922316_c19019e38f.jpg Attribution-ShareAlike License http://
creativecommons.org/licenses/by-sa/2.0/ 5468 6 c19019e38f 82de2f240e jpg 0
4436463882 42132616@N04 Miriam+Jones 2010-03-15 13:49:02.0 1268686142 San+Antonio
+House+Concert%2C+March+2010 http://www.flickr.com/photos/
42132616@N04/4436463882/ http://farm5.staticflickr.com/4007/4436463882_b96a3d9df9.jpg Attribution-
NoDerivs License http://creativecommons.org/licenses/by-nd/2.0/ 4007 5 b96a3d9df9
d901bc18a1 jpg 0
4572998878 78969707@N00 jkgreenstein12 2010-05-01 16:28:20.0 1272843440 Panasonic+DMC-LS70
Matt+and+Diana+Wedding+weekend Greeny%27s%2C+Christines%2C+Andy%2C+Jeeane+and+Lou%27s+photos+from
+the+wedding+weekend. diana,matt,wedding -71.047843 42.33719 16 http://
www.flickr.com/photos/78969707@N00/4572998878/ http://farm4.staticflickr.com/3373/4572998878_658b45226f.jpg
Attribution-NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/
3373 4 658b45226f a0777fdc0f jpg 0
9329902958 97468058@N00 angela+louise 2013-07-20 19:27:24.0 1374344844 Too+hot+%21
bantockpark,fayne,sean,strawberry,summer2013 http://
www.flickr.com/photos/97468058@N00/9329902958/ http://farm4.staticflickr.com/3819/9329902958_0bc80ce58a.jpg
Attribution-NonCommercial-ShareAlike License http://creativecommons.org/licenses/by-nc-sa/2.0/
3819 4 0bc80ce58a cd1d8e31ed jpg 0
3174965401 36813788@N00 Arthur+2+Sheds 2008-12-31 21:29:32.0 1231293771 Canon+EOS+DIGITAL
+REBEL+XTi IMG_3398 2009,new,years http://
www.flickr.com/photos/36813788@N00/3174965401/ http://farm4.staticflickr.com/3099/3174965401_2742004421.jpg
Attribution-NonCommercial-NoDerivs License http://creativecommons.org/licenses/by-nc-nd/2.0/
3099 4 2742004421 486c7d19b5 jpg 0

```

Acquiring Data: YFCC100M Dataset

100 Million Photos + Videos

- Photo/video identifier
- User NSID, User nickname
- **Date taken**
- Date uploaded
- **Capture device**
- Title, Description
- User tags (comma-separated)
- Machine tags (comma-separated)
- **Longitude**
- **Latitude**
- Accuracy
- Photo/video page URL, Photo/video download URL
- License name, License URL
- Photo/video server identifier, Photo/video farm identifier
- Photo/video secret, Photo/video secret original
- Photo/video extension original
- Photos/video marker (0 = photo, 1 = video)

```

# Taking out scanners using wordinentry()
wordnotinentry =
np.invert(wordinentry(flickr_df,'camera','scan'))
flickr_df = flickr_df[wordnotinentry]

# Create a new column for camera brand
flickr_df['camera_brand'] =
flickr_df.camera.map(lambda x: x.split('+')[0])

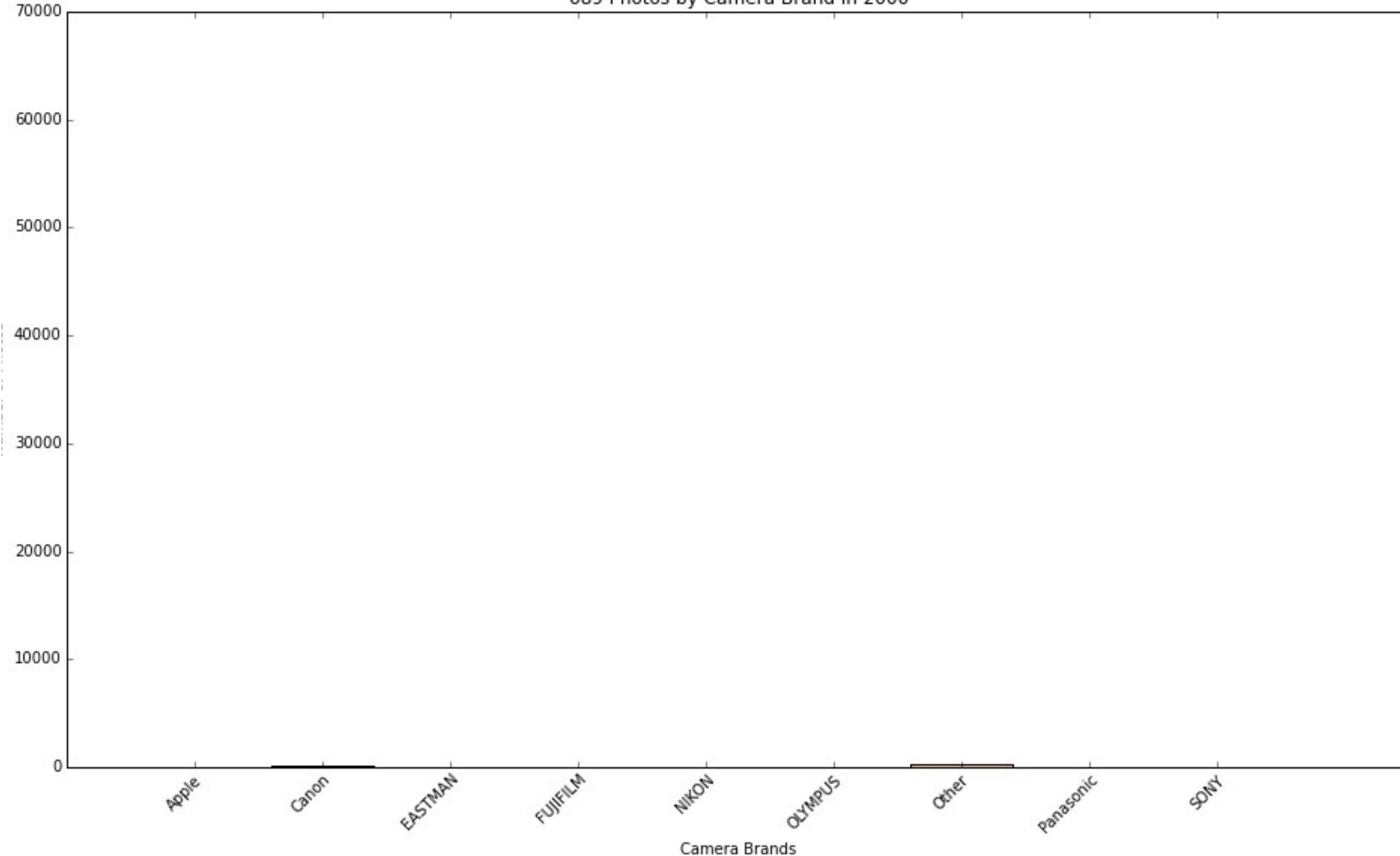
# Binning brands into "Other" if frequency < 0.01
flickr_df = cleanup_data(flickr_df,'camera_brand',
0.01)

```

Cleaning Data

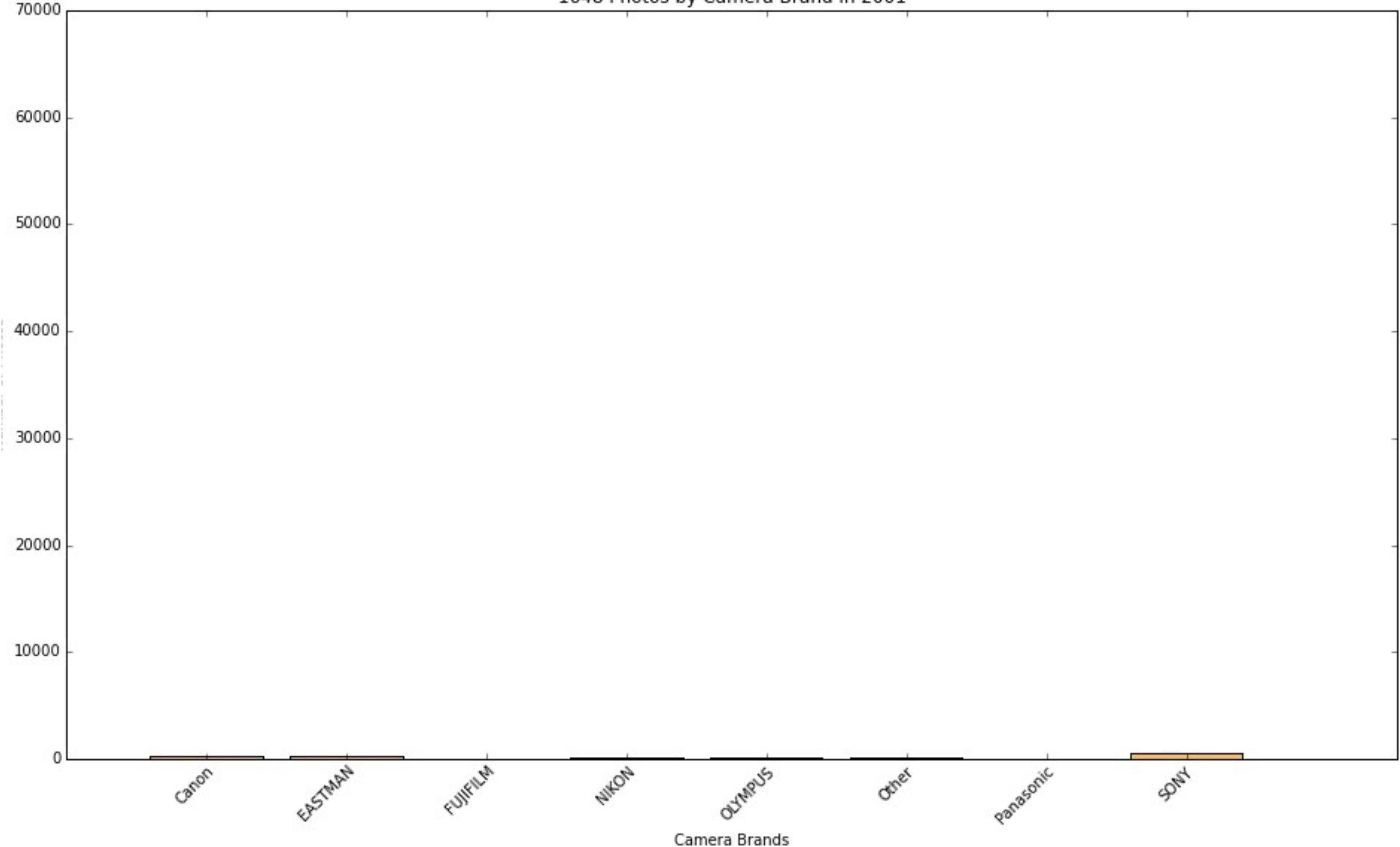
Preprocessing Text

689 Photos by Camera Brand in 2000



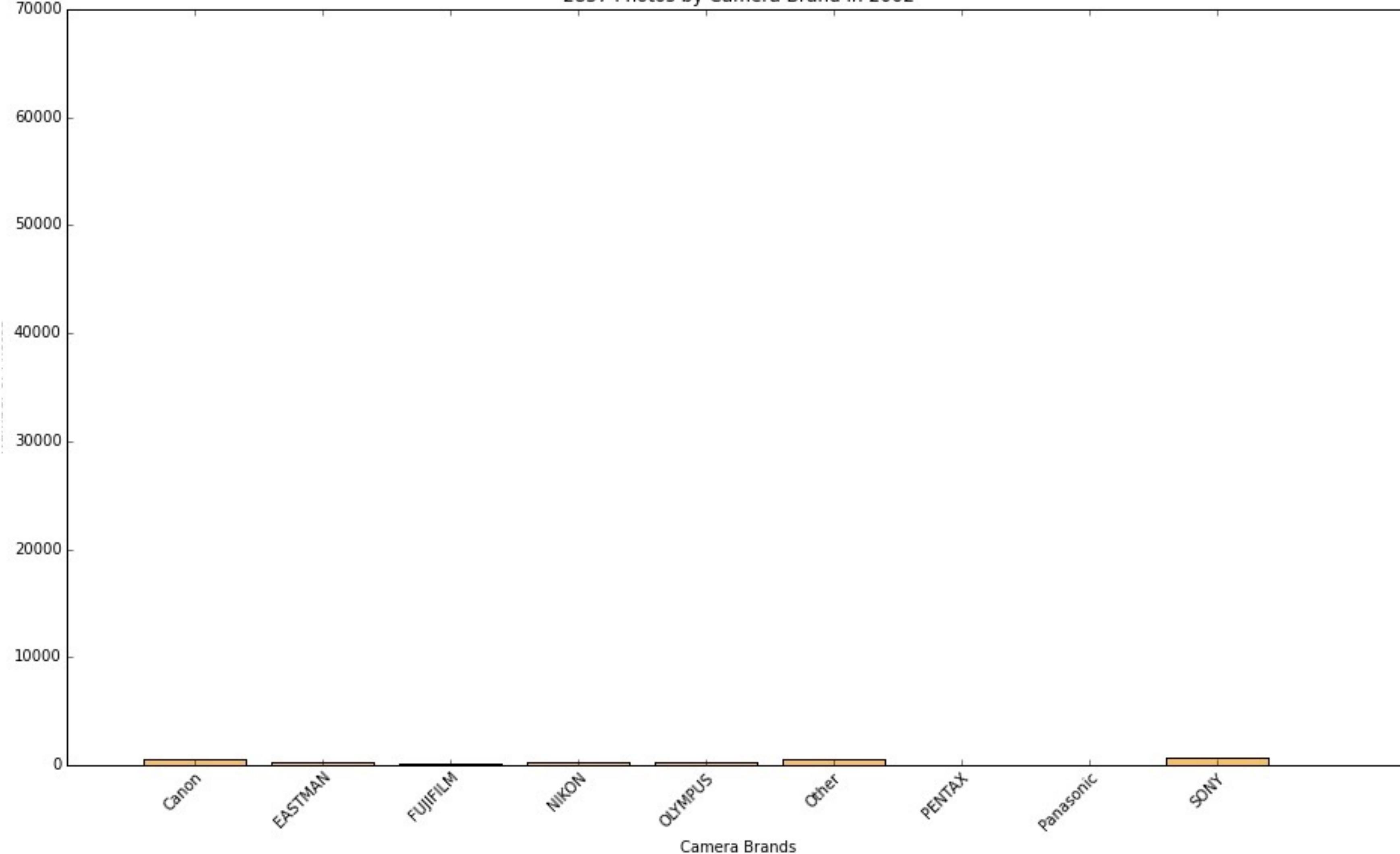
Visual Explorations
Photos by Camera Brand

1648 Photos by Camera Brand in 2001



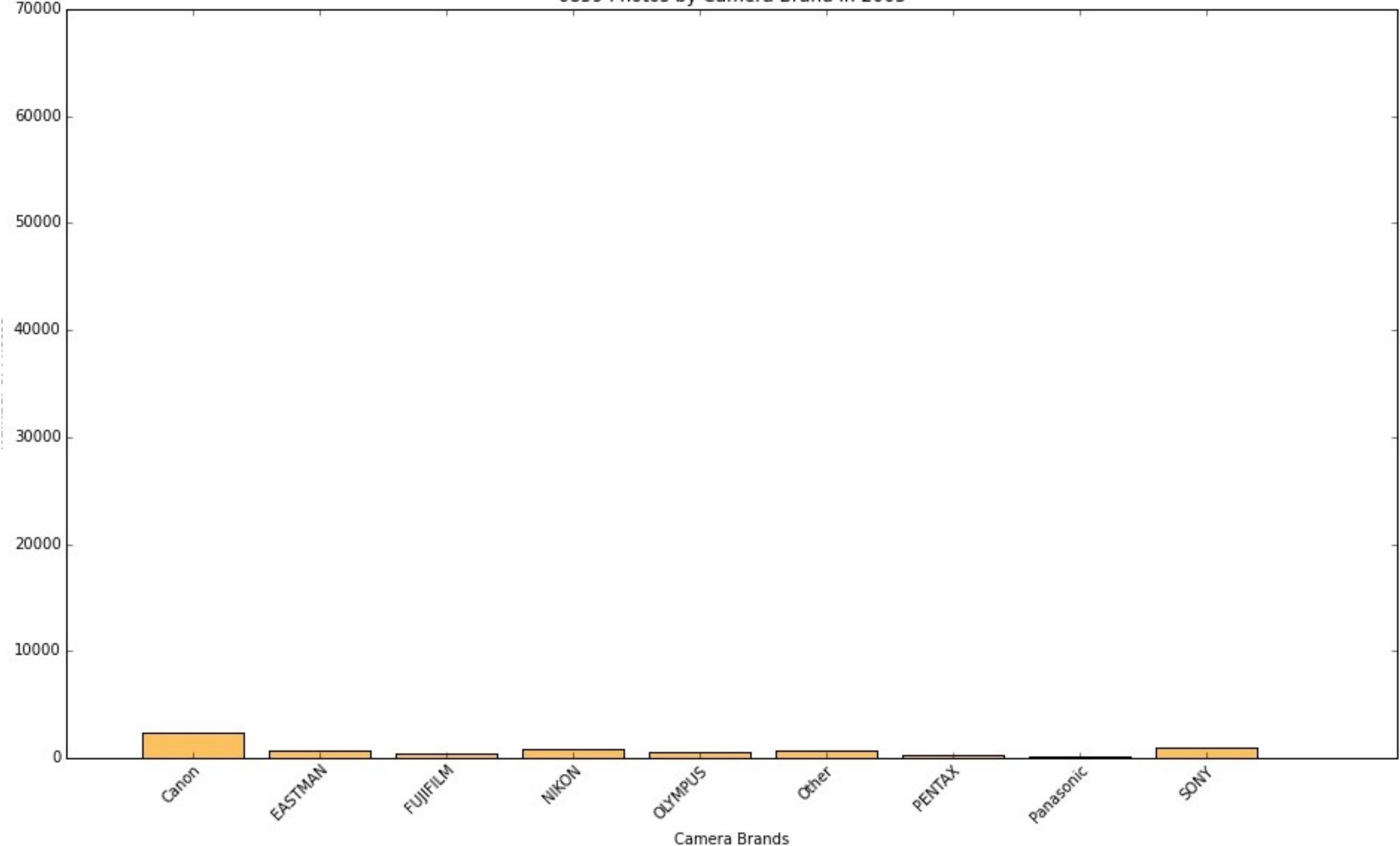
Visual Explorations
Photos by Camera Brand

2837 Photos by Camera Brand in 2002



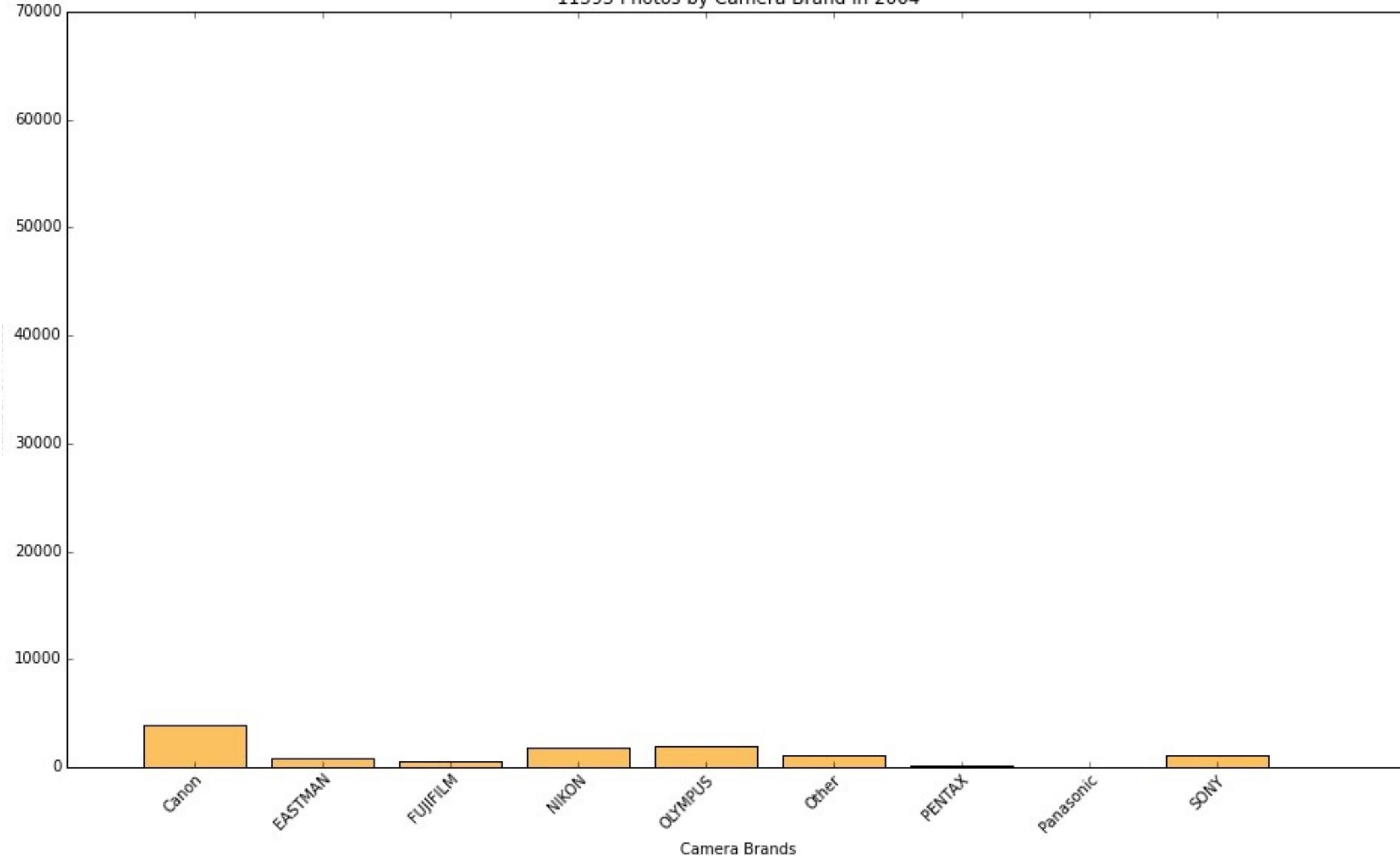
Visual Explorations
Photos by Camera Brand

6859 Photos by Camera Brand in 2003



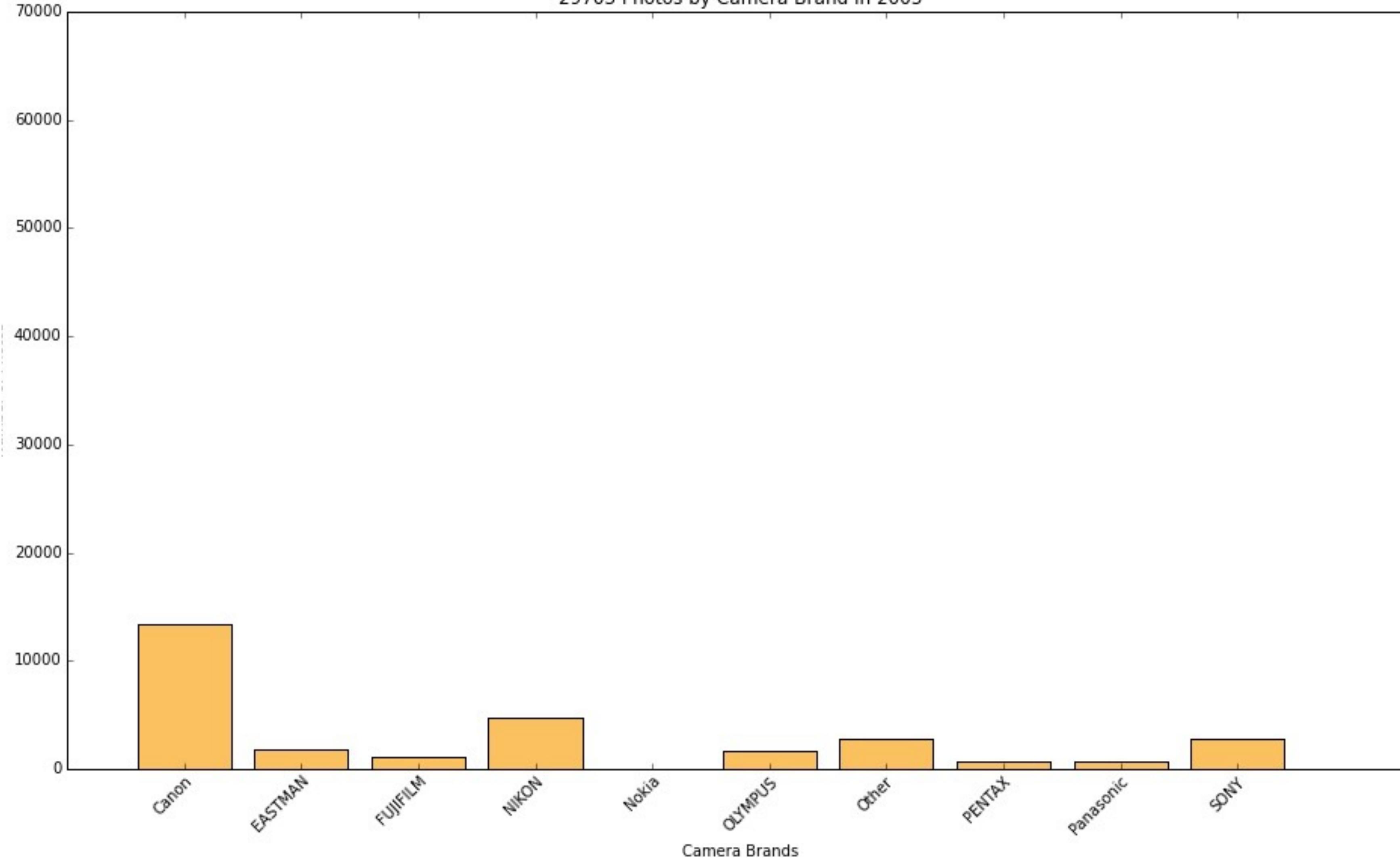
Visual Explorations
Photos by Camera Brand

11393 Photos by Camera Brand in 2004



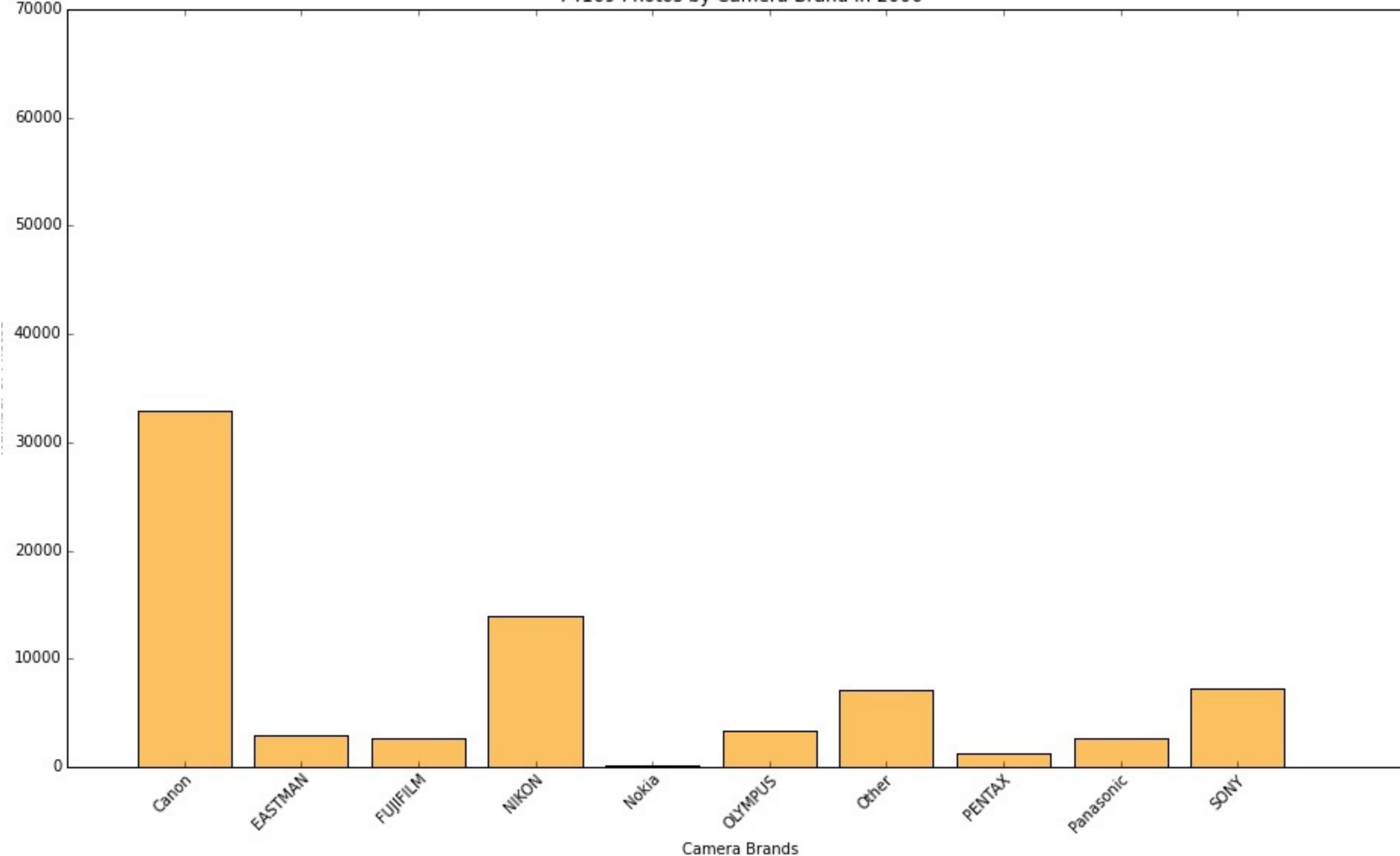
Visual Explorations
Photos by Camera Brand

29703 Photos by Camera Brand in 2005



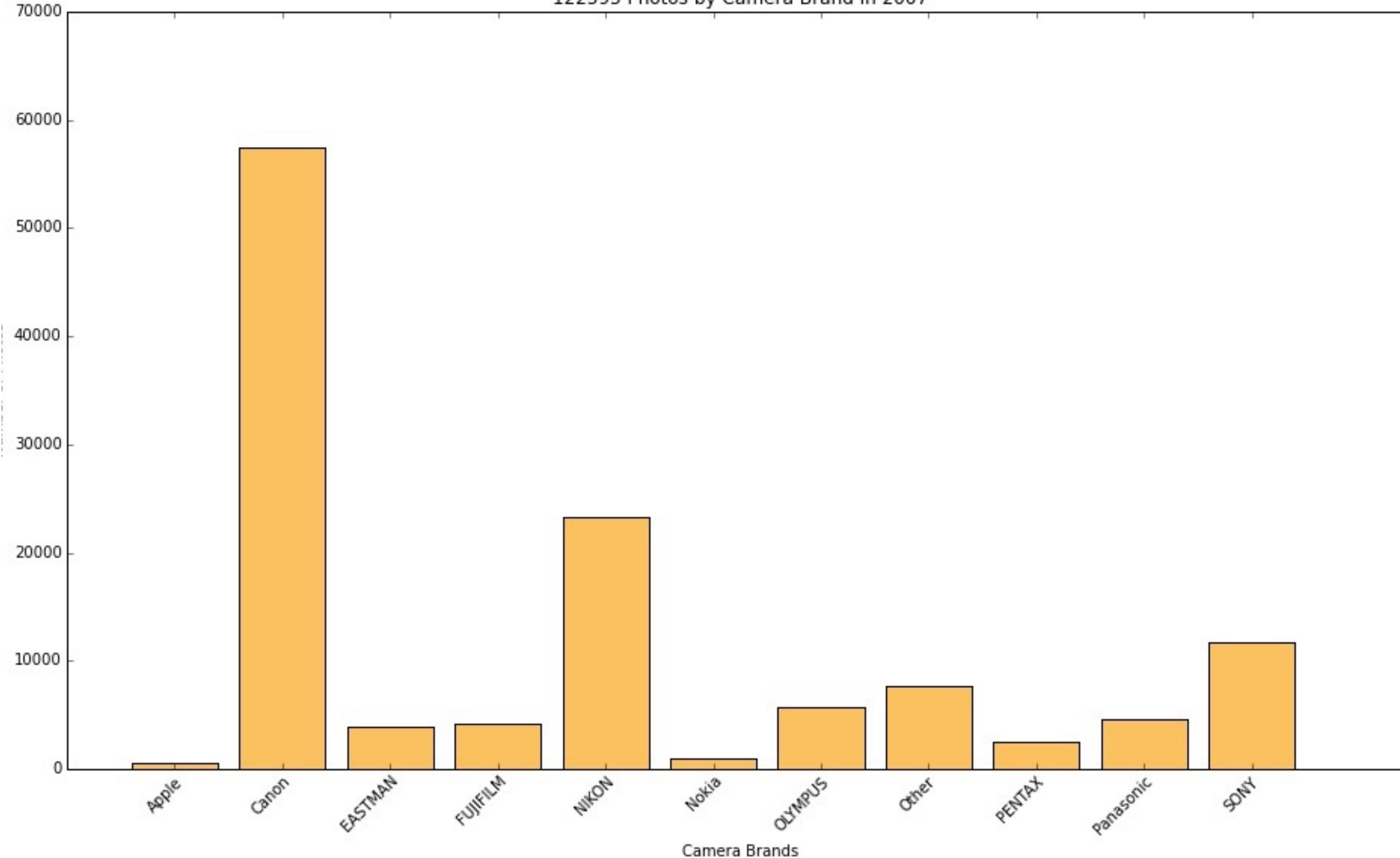
Visual Explorations
Photos by Camera Brand

74169 Photos by Camera Brand in 2006



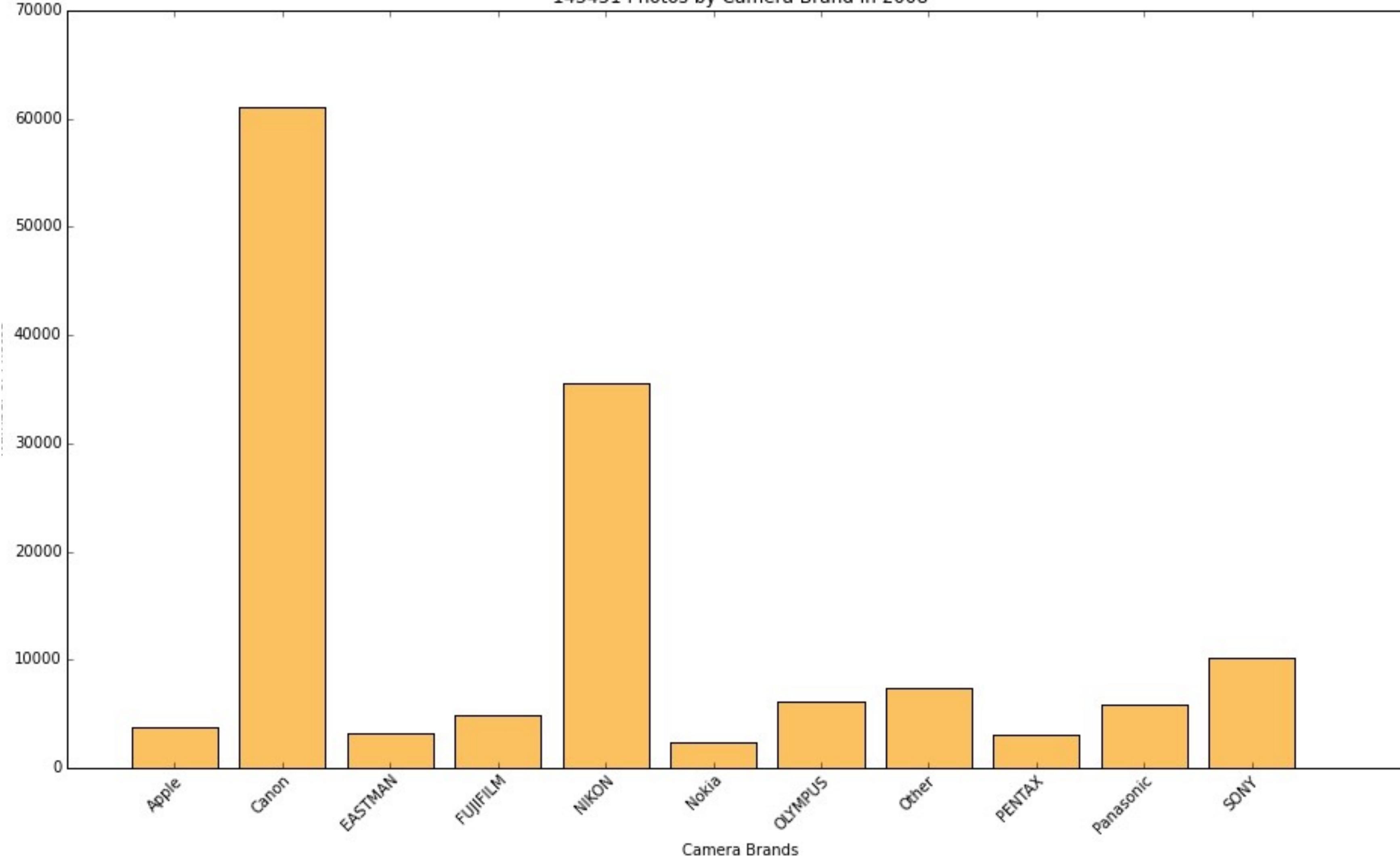
Visual Explorations
Photos by Camera Brand

122595 Photos by Camera Brand in 2007



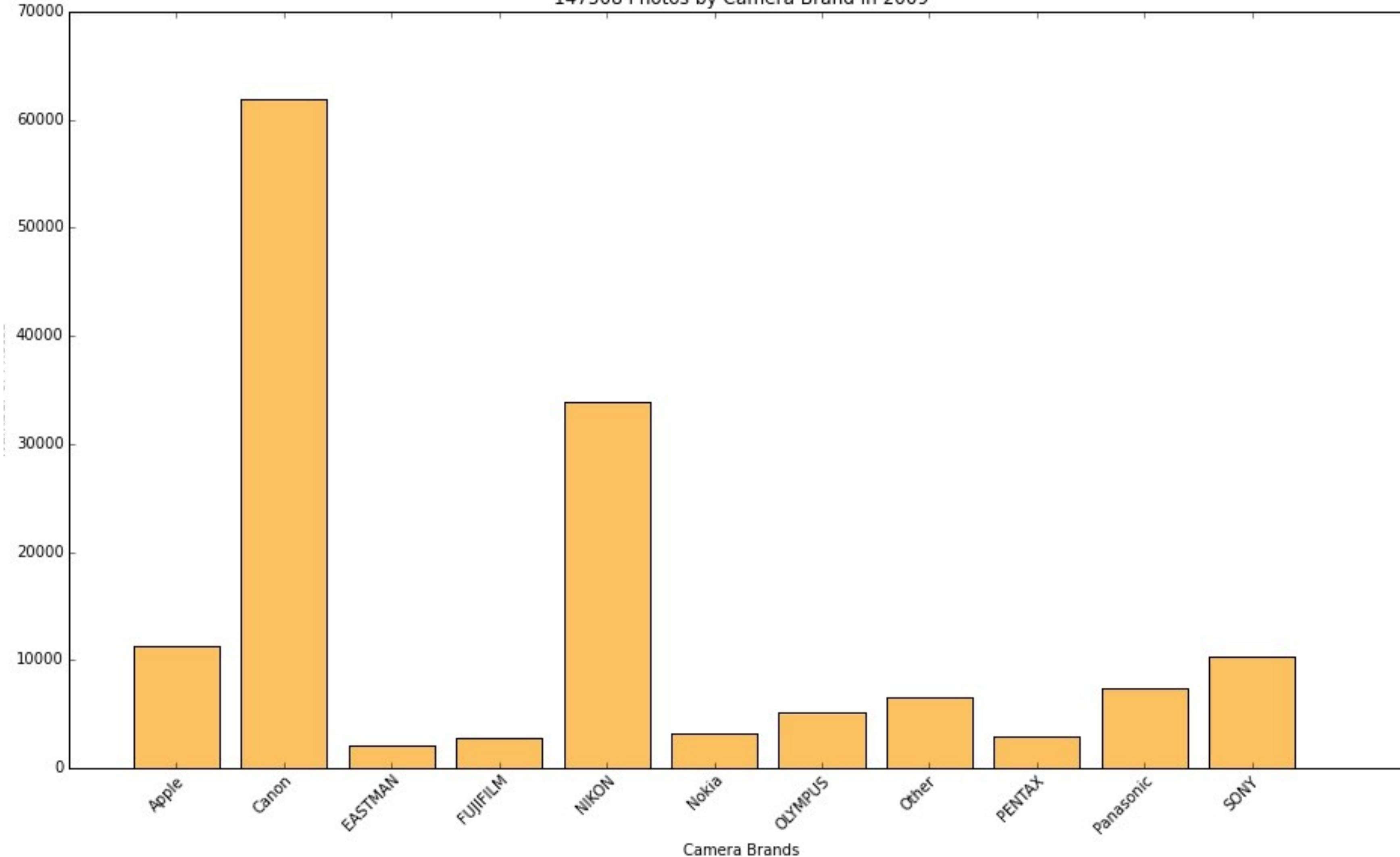
Visual Explorations
Photos by Camera Brand

143451 Photos by Camera Brand in 2008



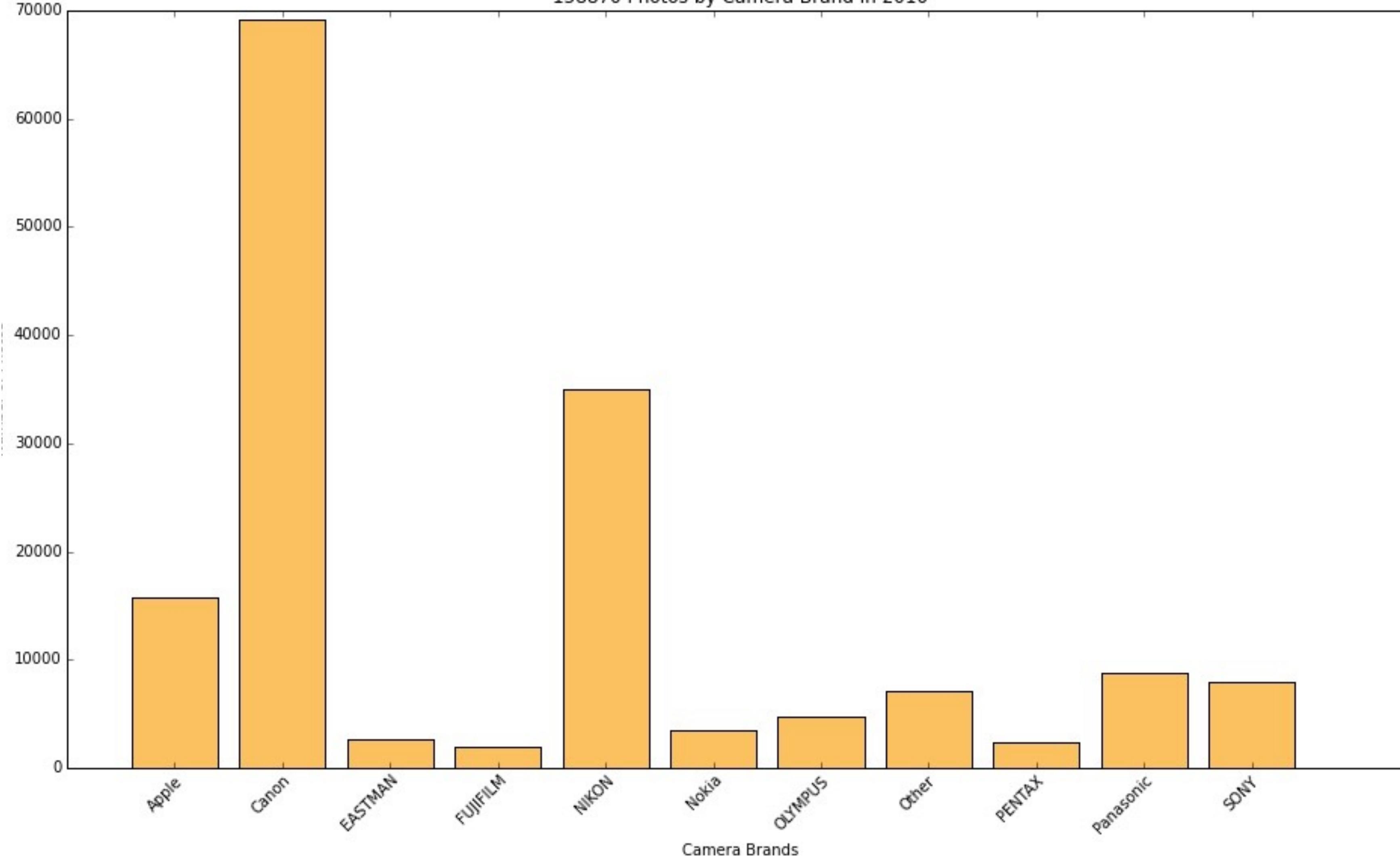
Visual Explorations
Photos by Camera Brand

147508 Photos by Camera Brand in 2009



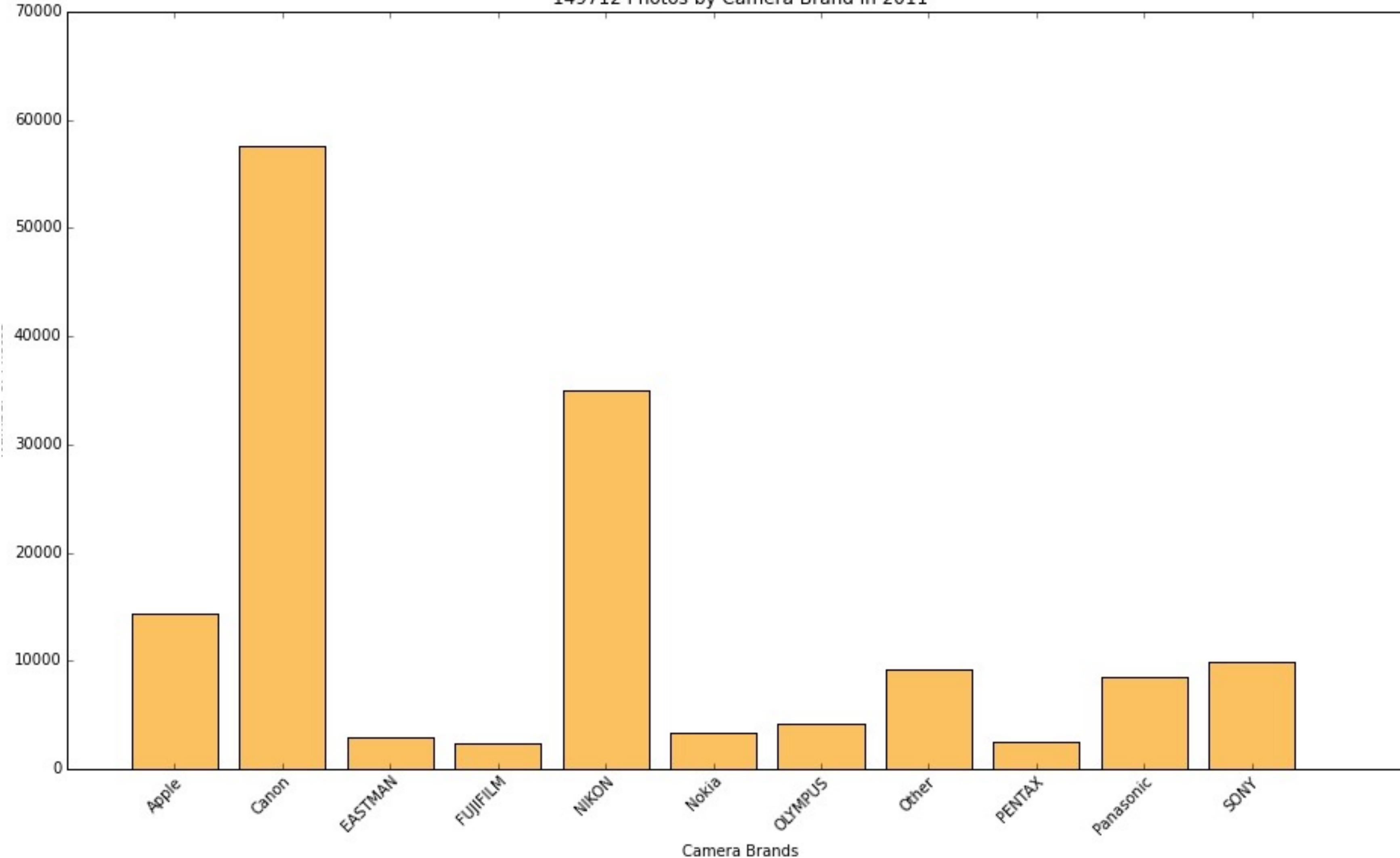
Visual Explorations
Photos by Camera Brand

158870 Photos by Camera Brand in 2010



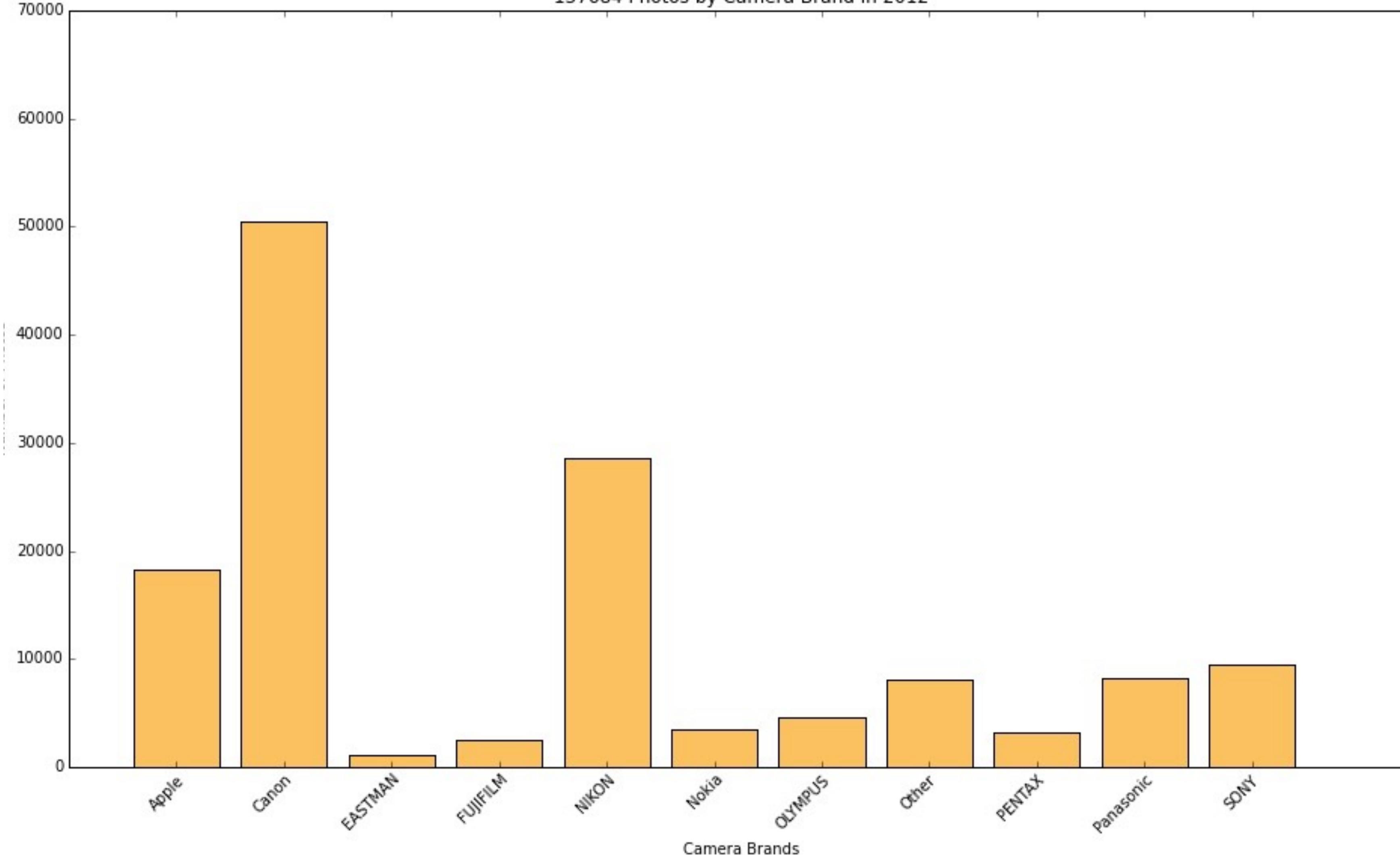
Visual Explorations
Photos by Camera Brand

149712 Photos by Camera Brand in 2011



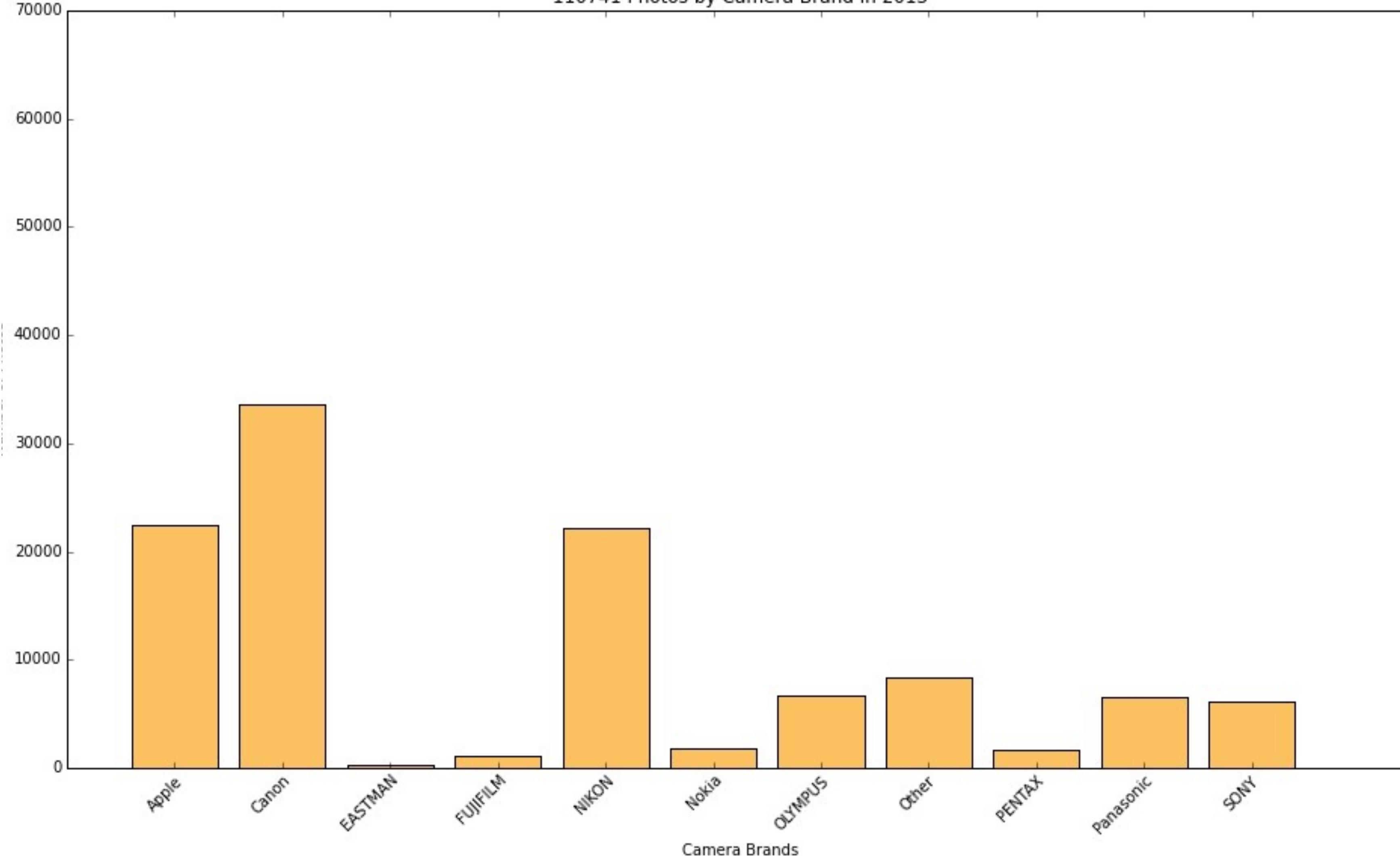
Visual Explorations
Photos by Camera Brand

137684 Photos by Camera Brand in 2012



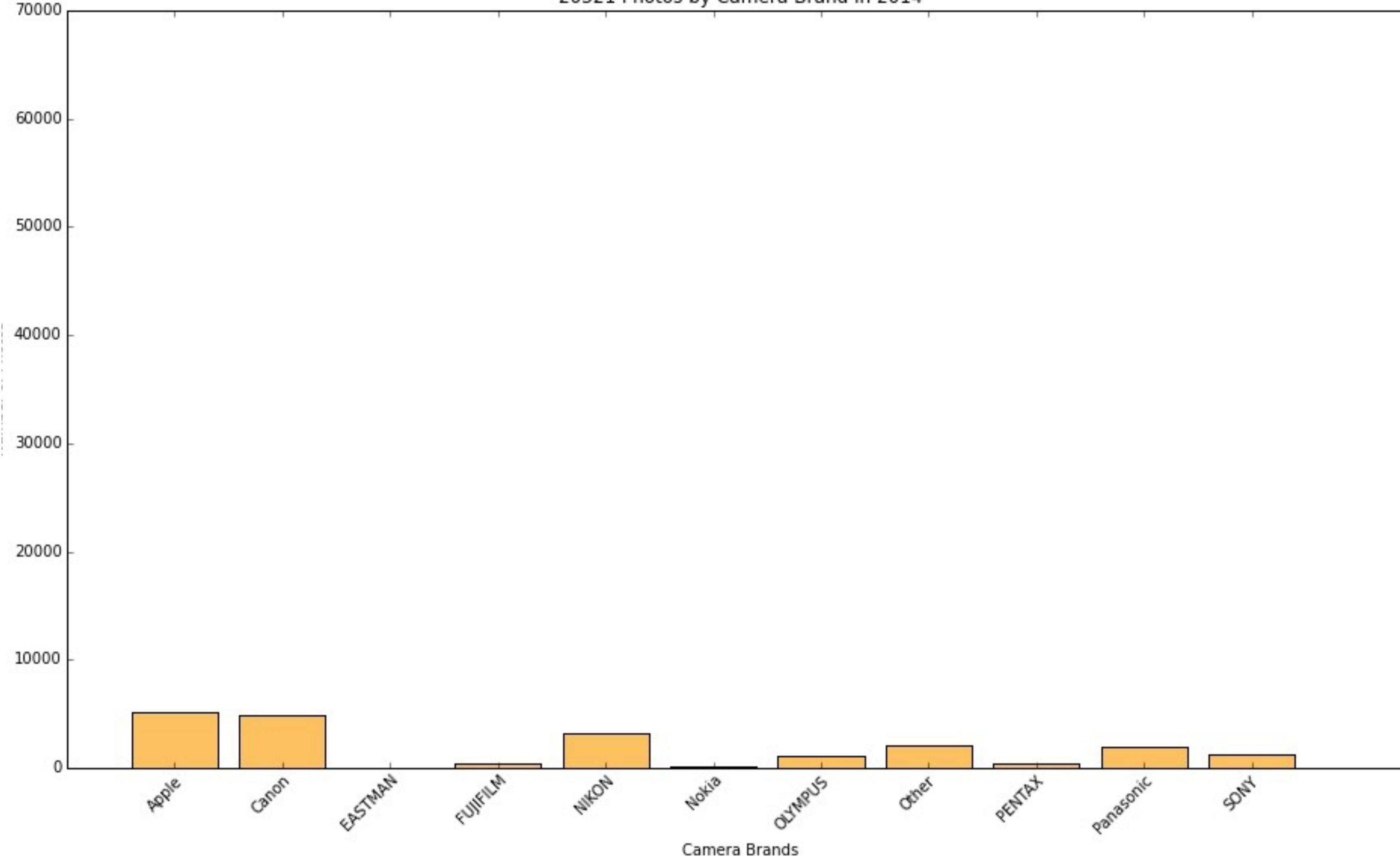
Visual Explorations
Photos by Camera Brand

110741 Photos by Camera Brand in 2013

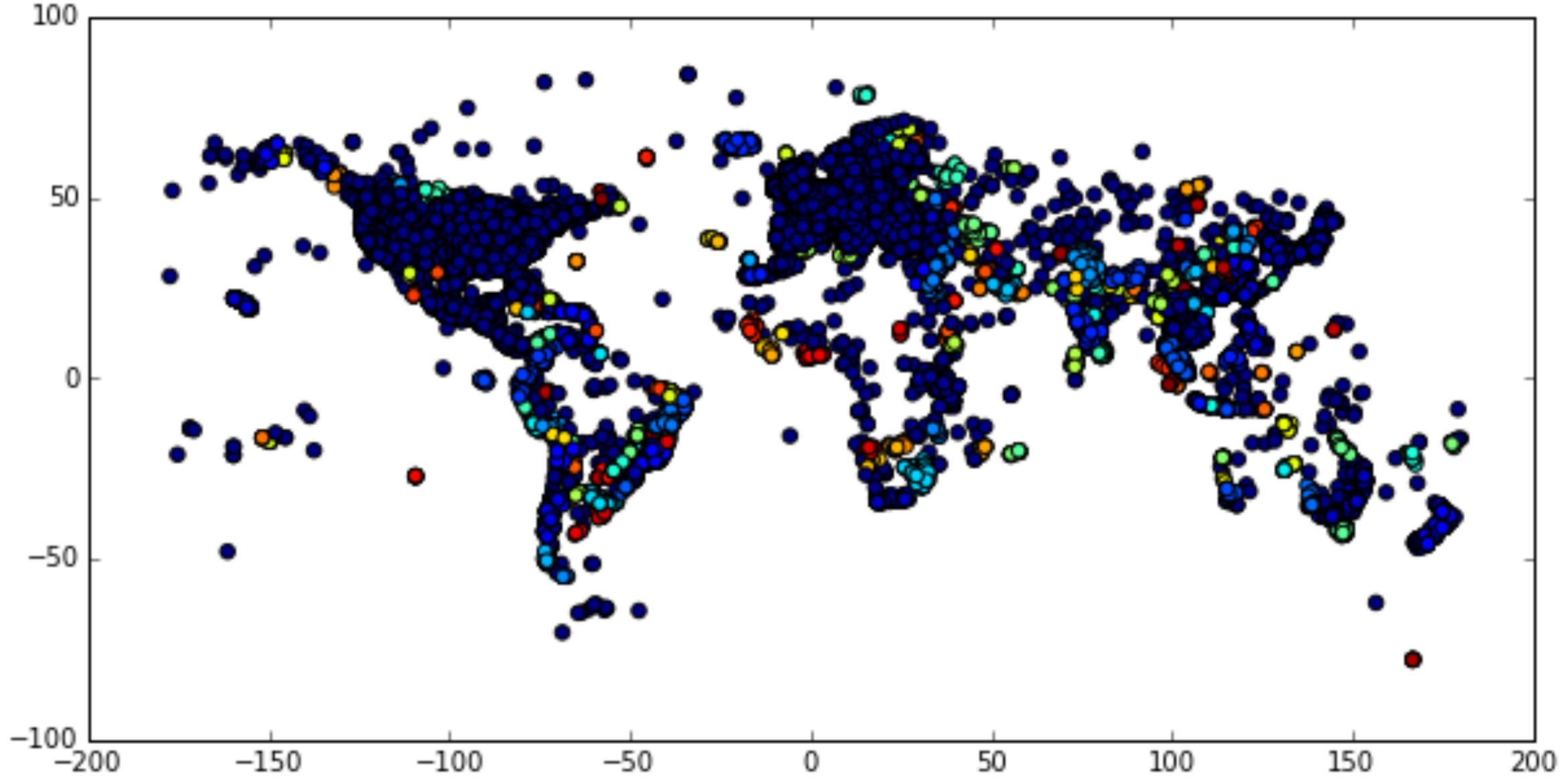


Visual Explorations
Photos by Camera Brand

20321 Photos by Camera Brand in 2014

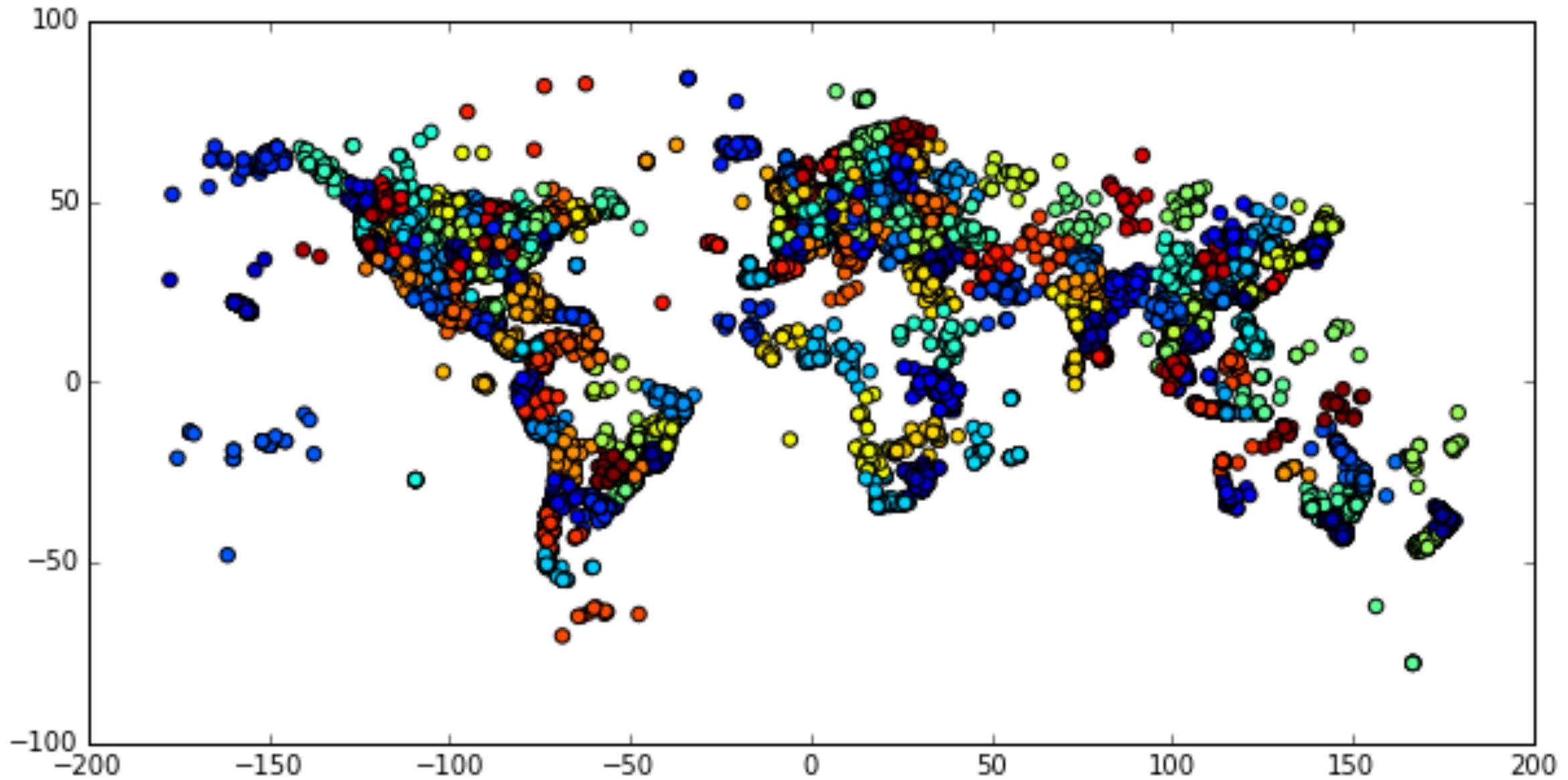


Visual Explorations
Photos by Camera Brand



DBSCAN

- Better for capturing outliers
- Most populous places are categorized together
 - North America/ Europe

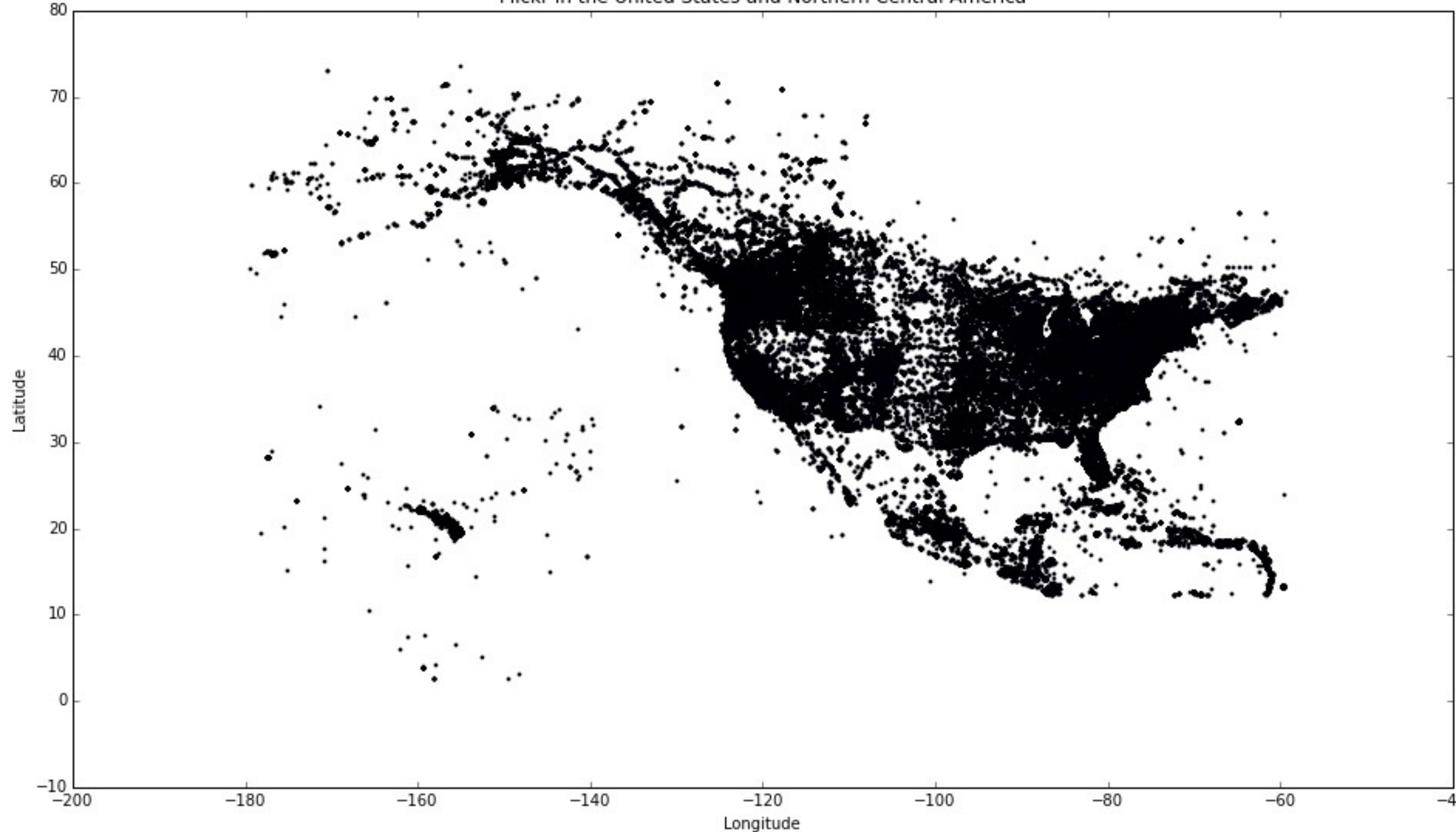


K-Means

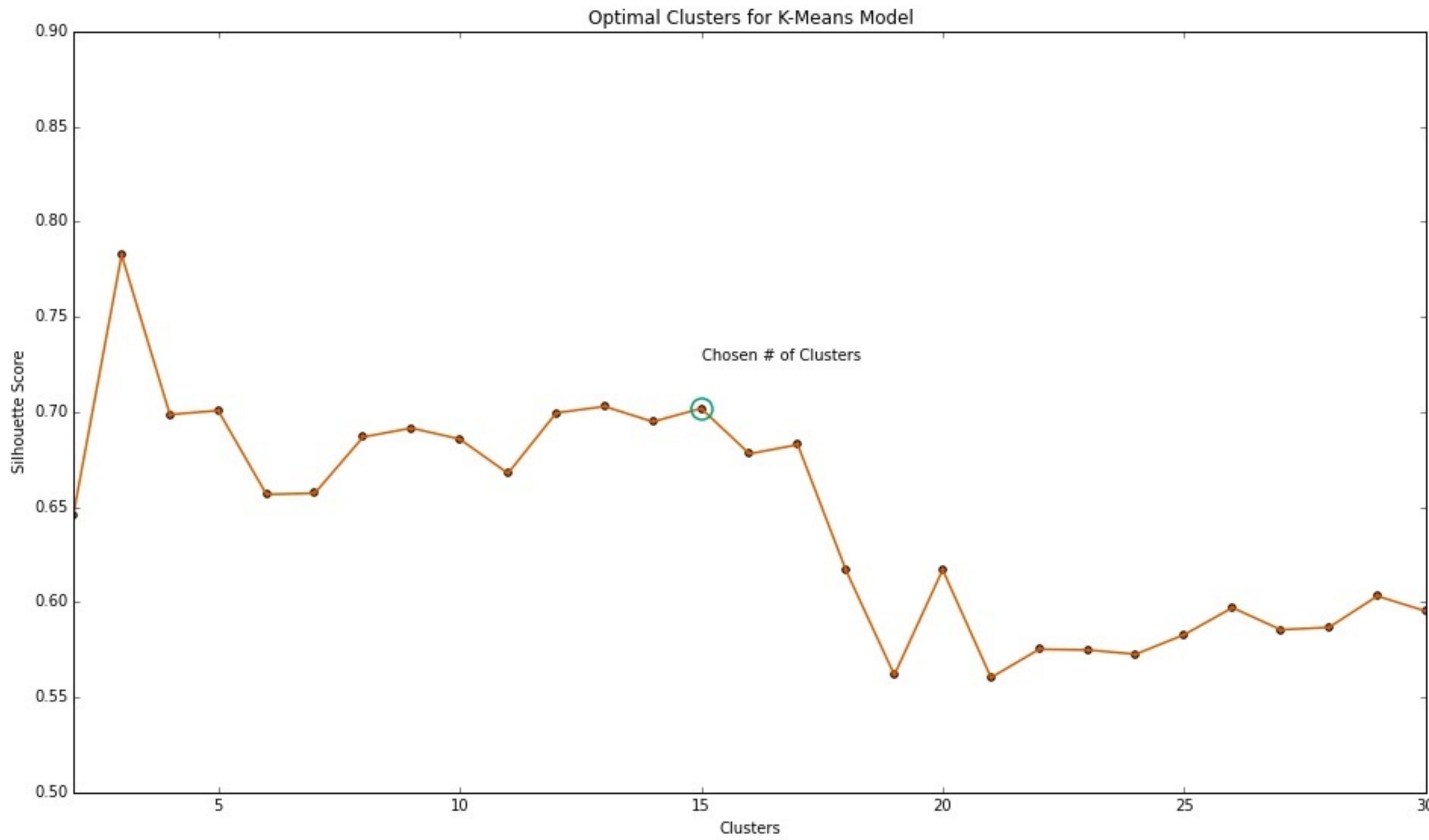
- Needs too many regions to define

Visual Explorations
Clustering Methods

Flickr in the United States and Northern Central America



Visual Explorations
Photos in the United States



Silhouette Score

Used a sample size of 10,000 for computation

Tested range between 1 and 30 clusters

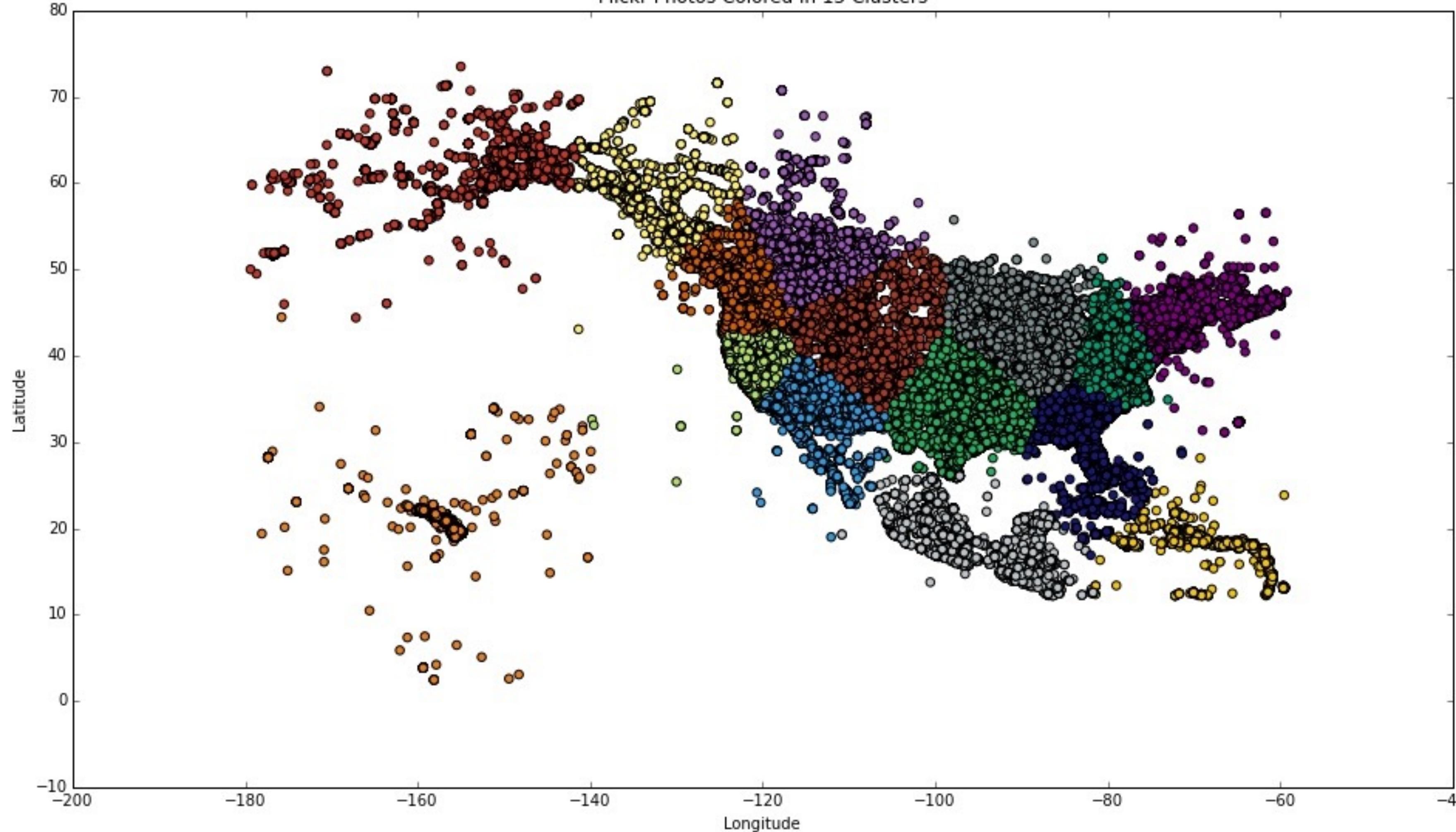
...
11: 0.668
12: 0.699
13: 0.702
14: 0.695
15: 0.702
16: 0.678
17: 0.683
18: 0.617
...

Final: **15 clusters**

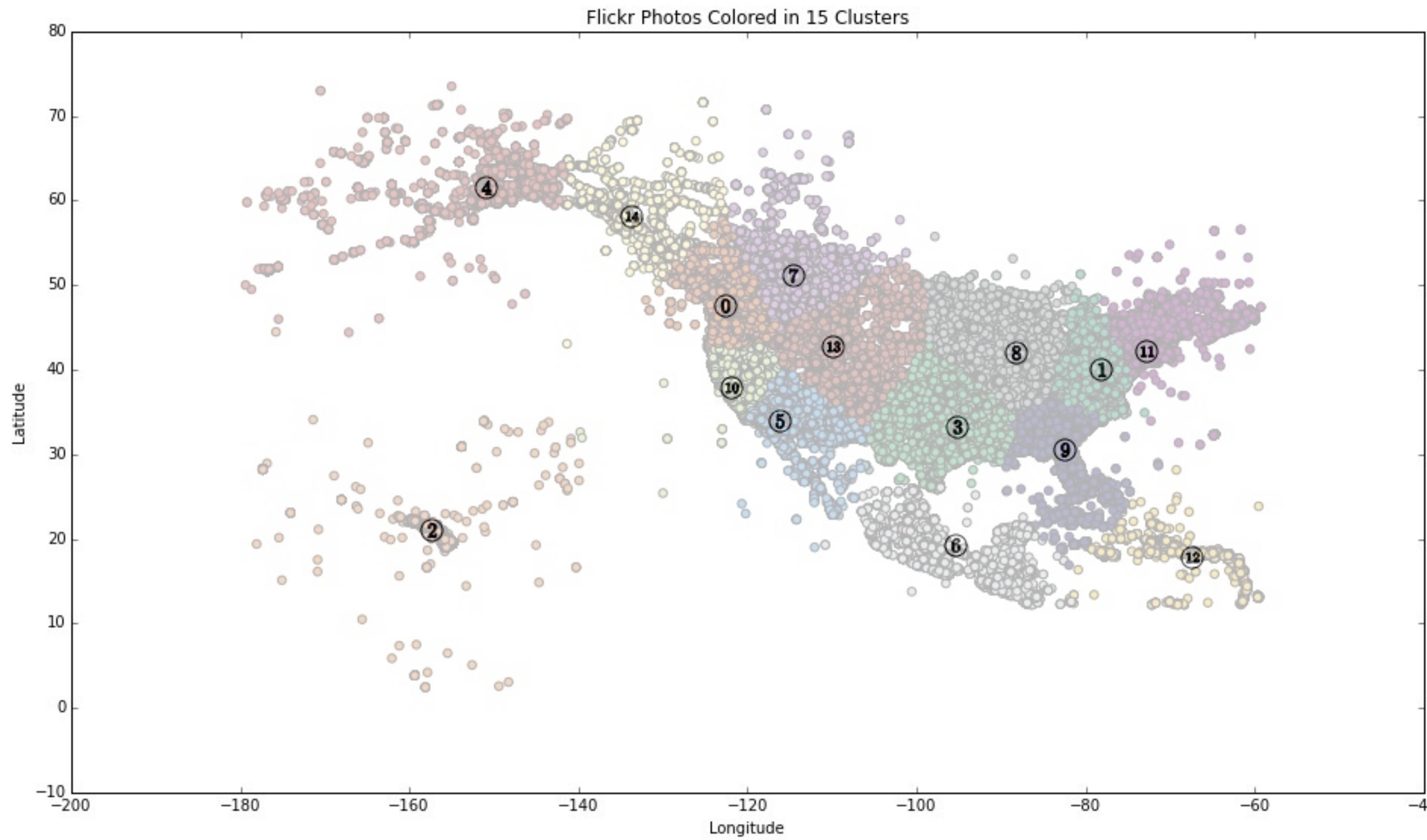
K-Means Clusters

Clustering into Regions

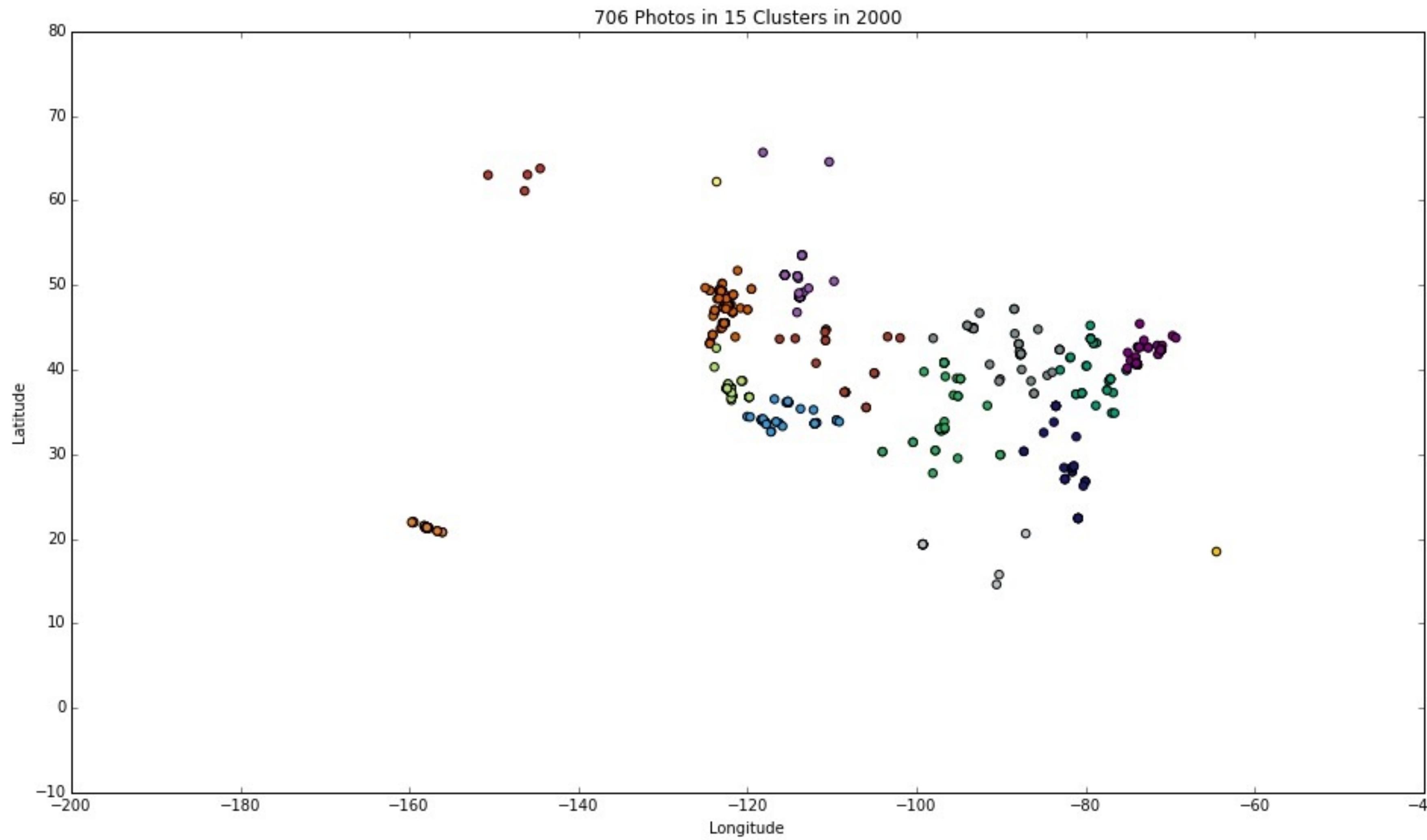
Flickr Photos Colored in 15 Clusters



K-Means Clusters
Clustering into Regions

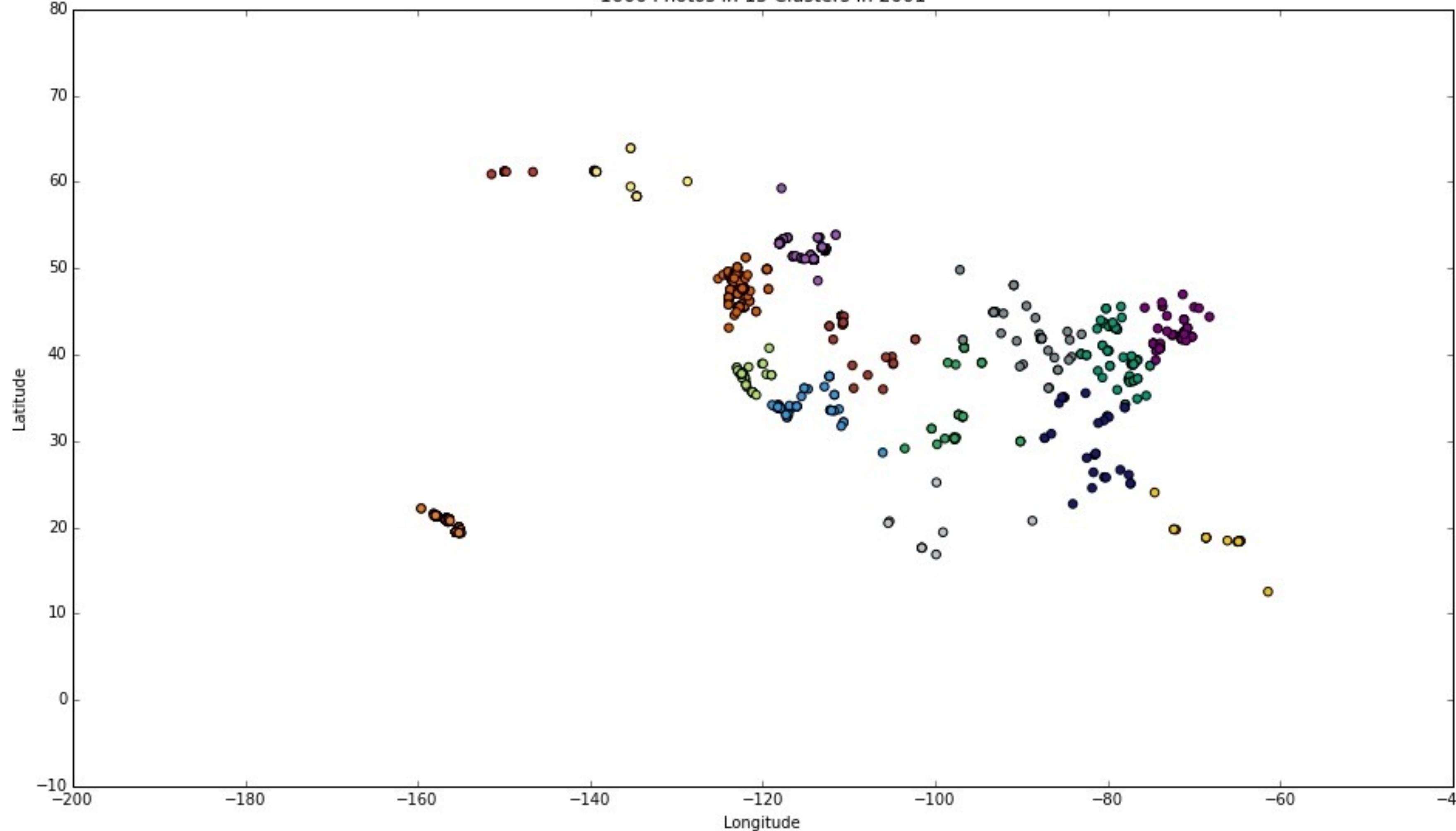


K-Means Clusters
Clustering into Regions



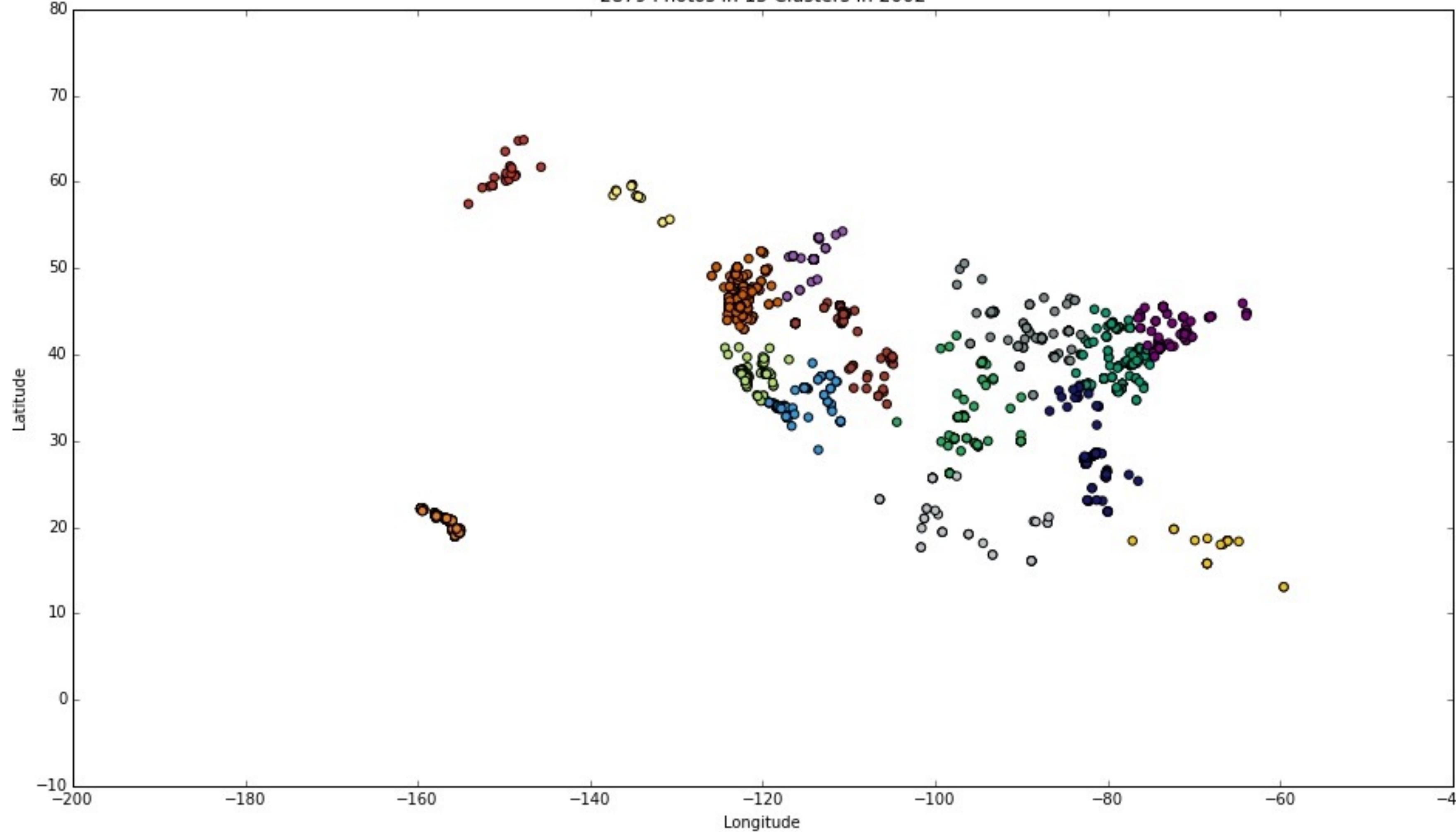
K-Means Clusters
Cluster Regions by Year

1666 Photos in 15 Clusters in 2001



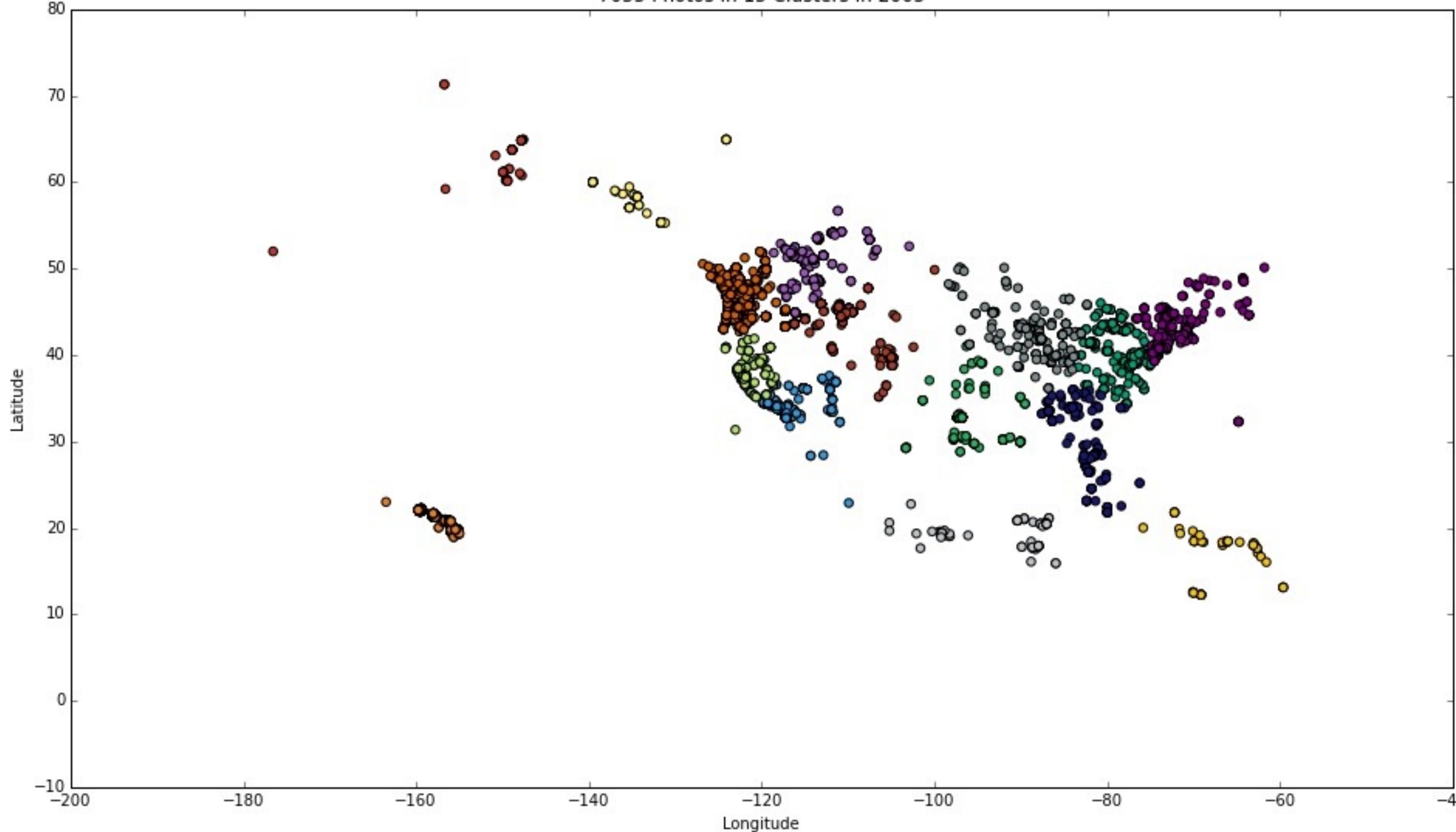
K-Means Clusters
Cluster Regions by Year

2879 Photos in 15 Clusters in 2002



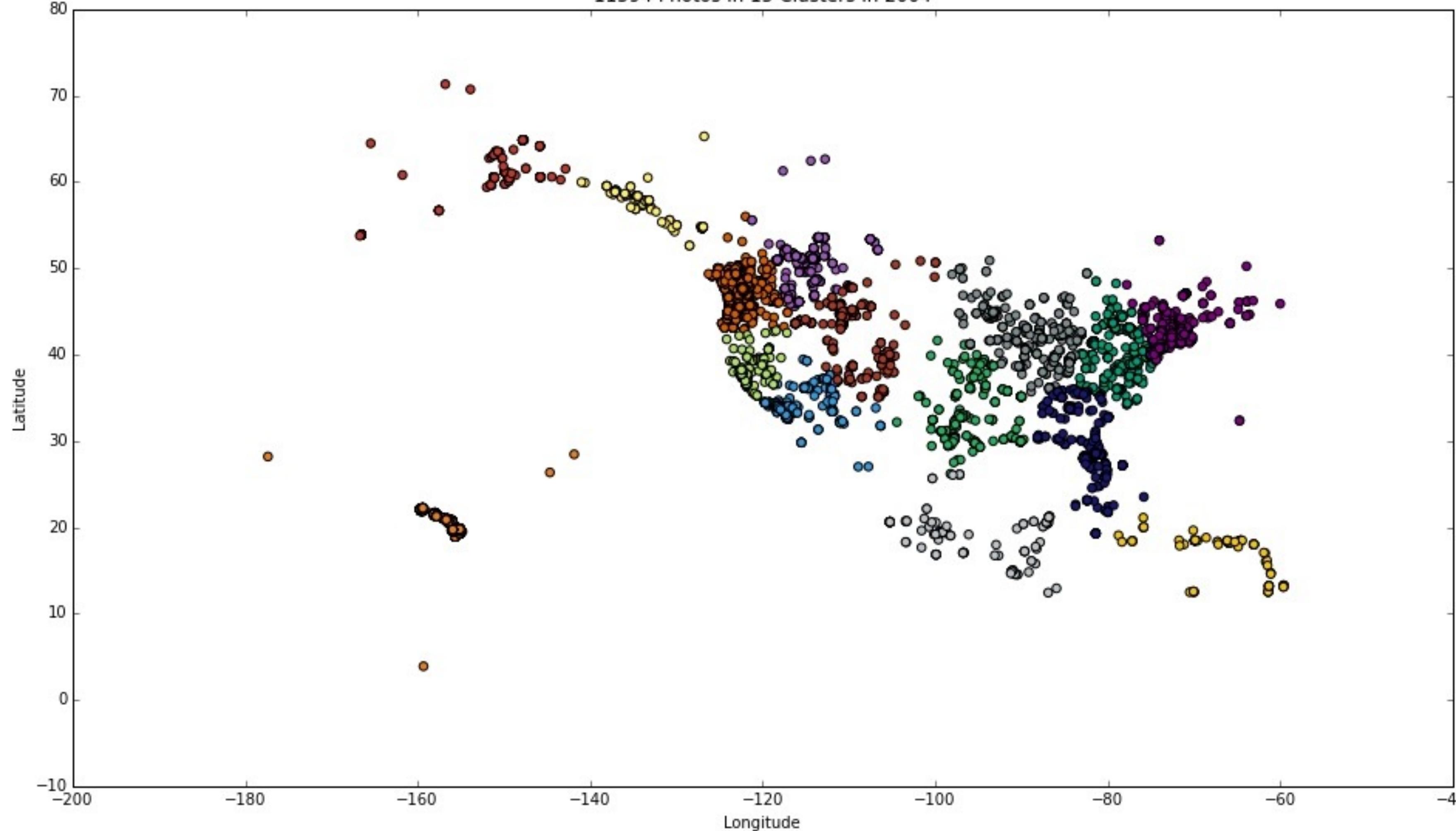
K-Means Clusters
Cluster Regions by Year

7055 Photos in 15 Clusters in 2003



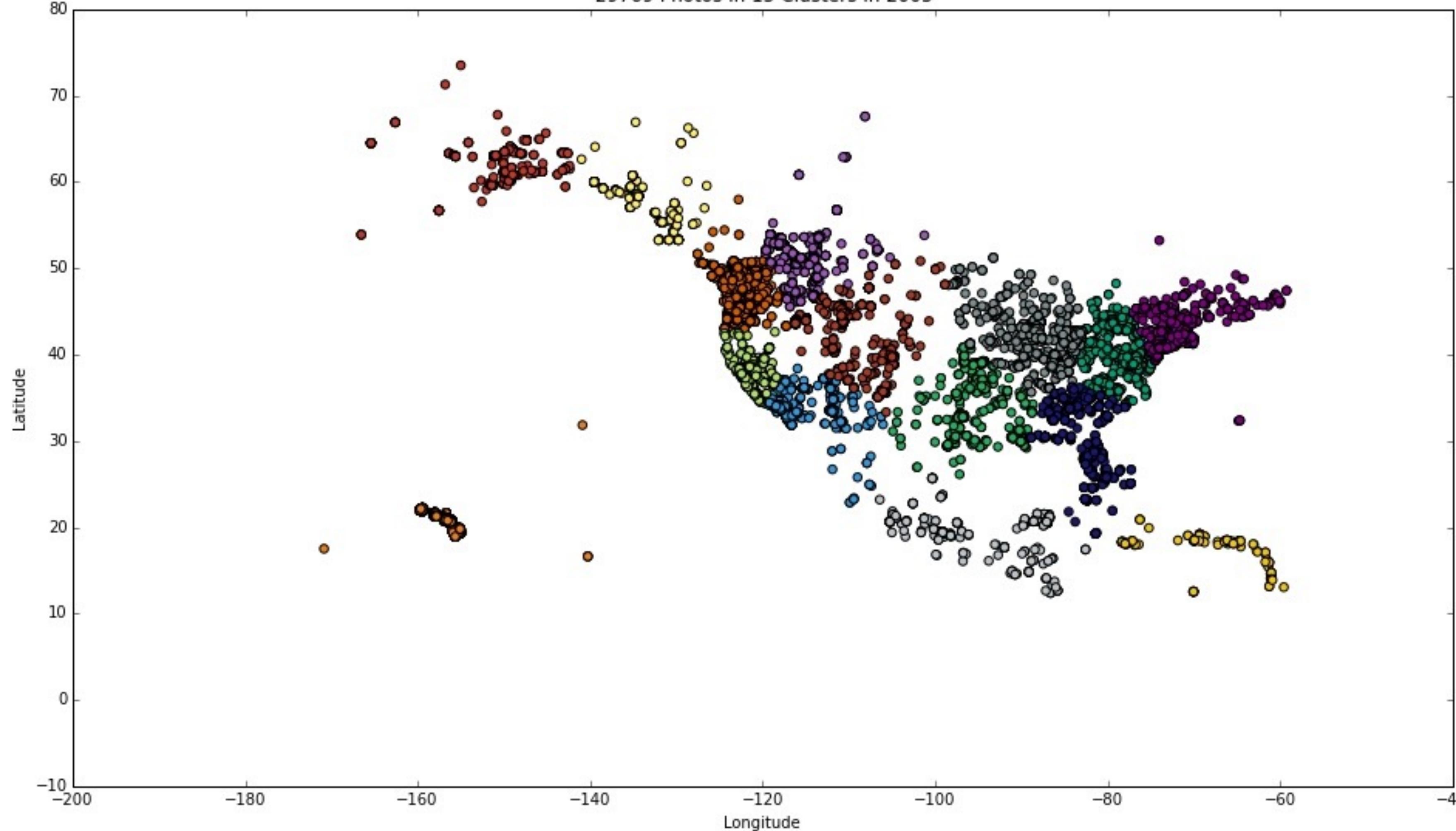
K-Means Clusters
Cluster Regions by Year

11594 Photos in 15 Clusters in 2004



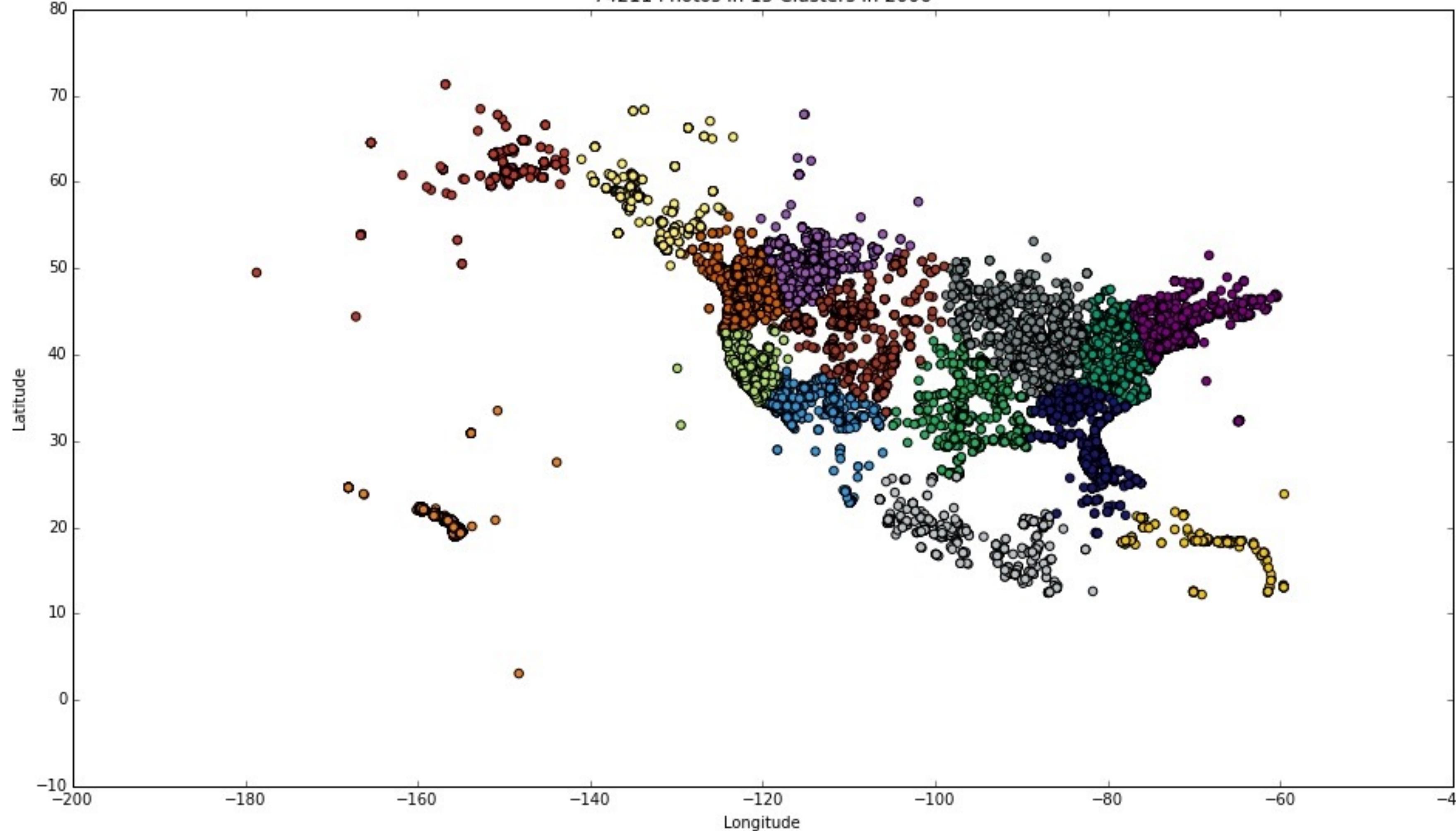
K-Means Clusters
Cluster Regions by Year

29769 Photos in 15 Clusters in 2005



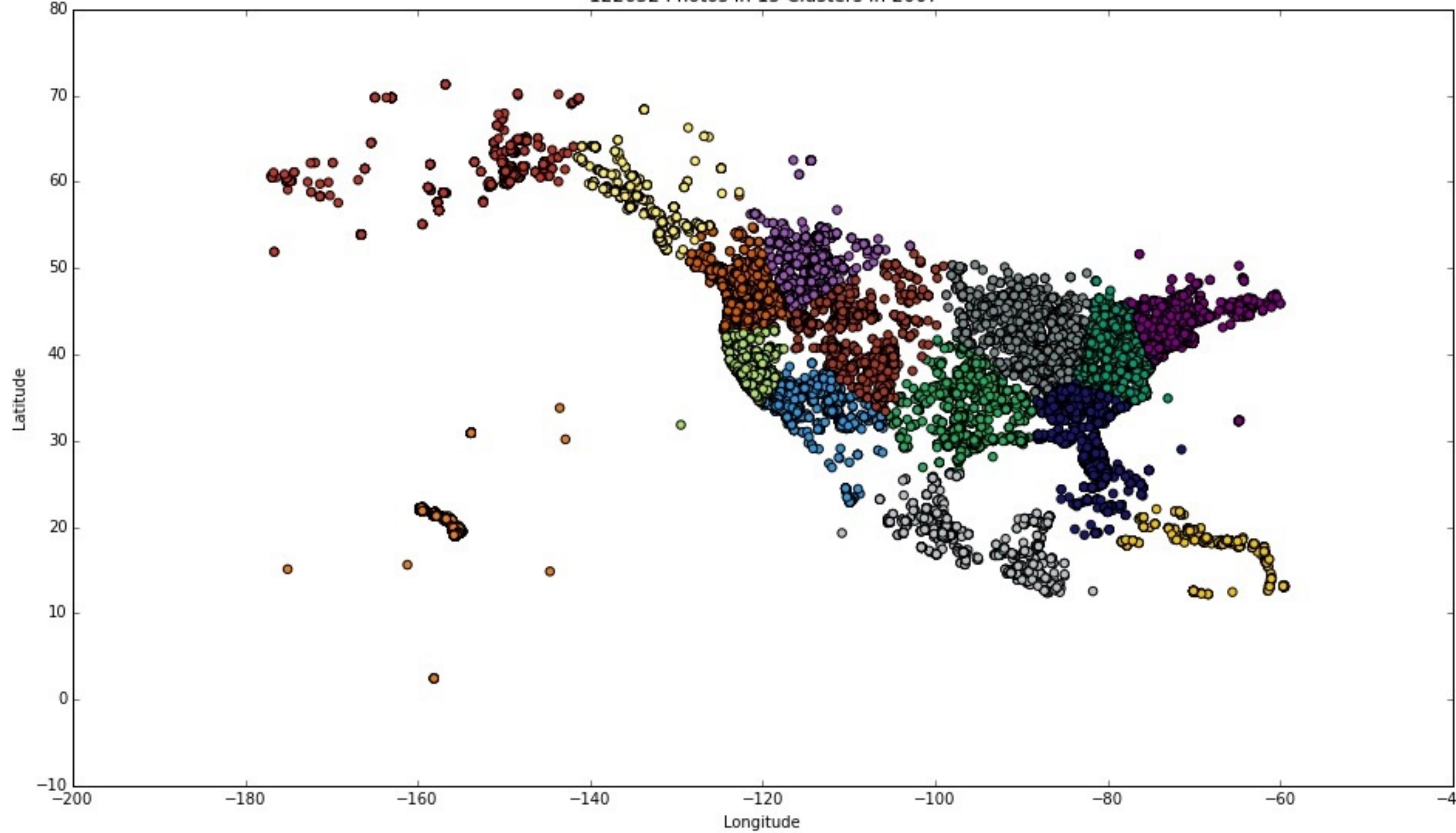
K-Means Clusters
Cluster Regions by Year

74211 Photos in 15 Clusters in 2006



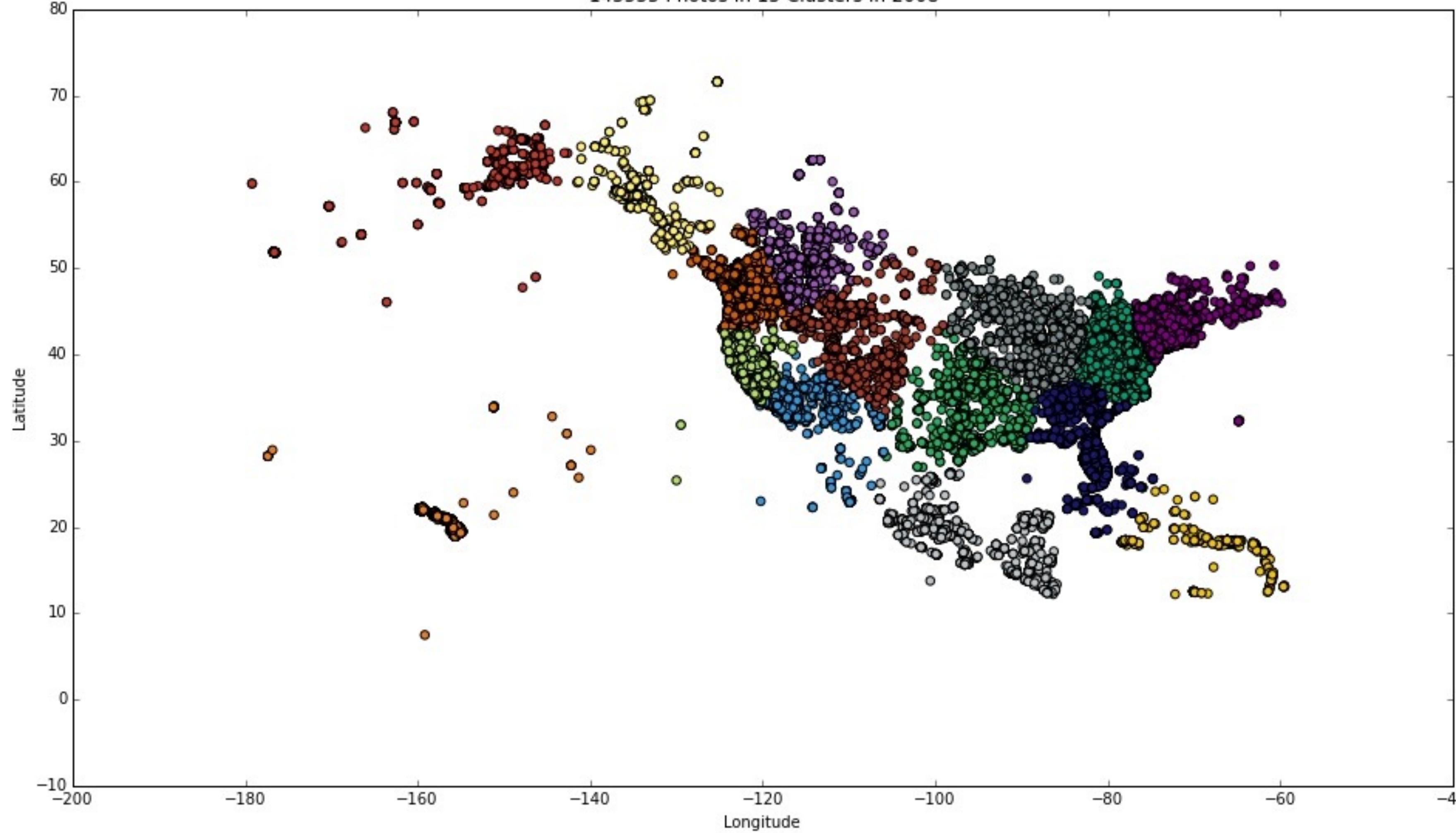
K-Means Clusters
Cluster Regions by Year

122632 Photos in 15 Clusters in 2007



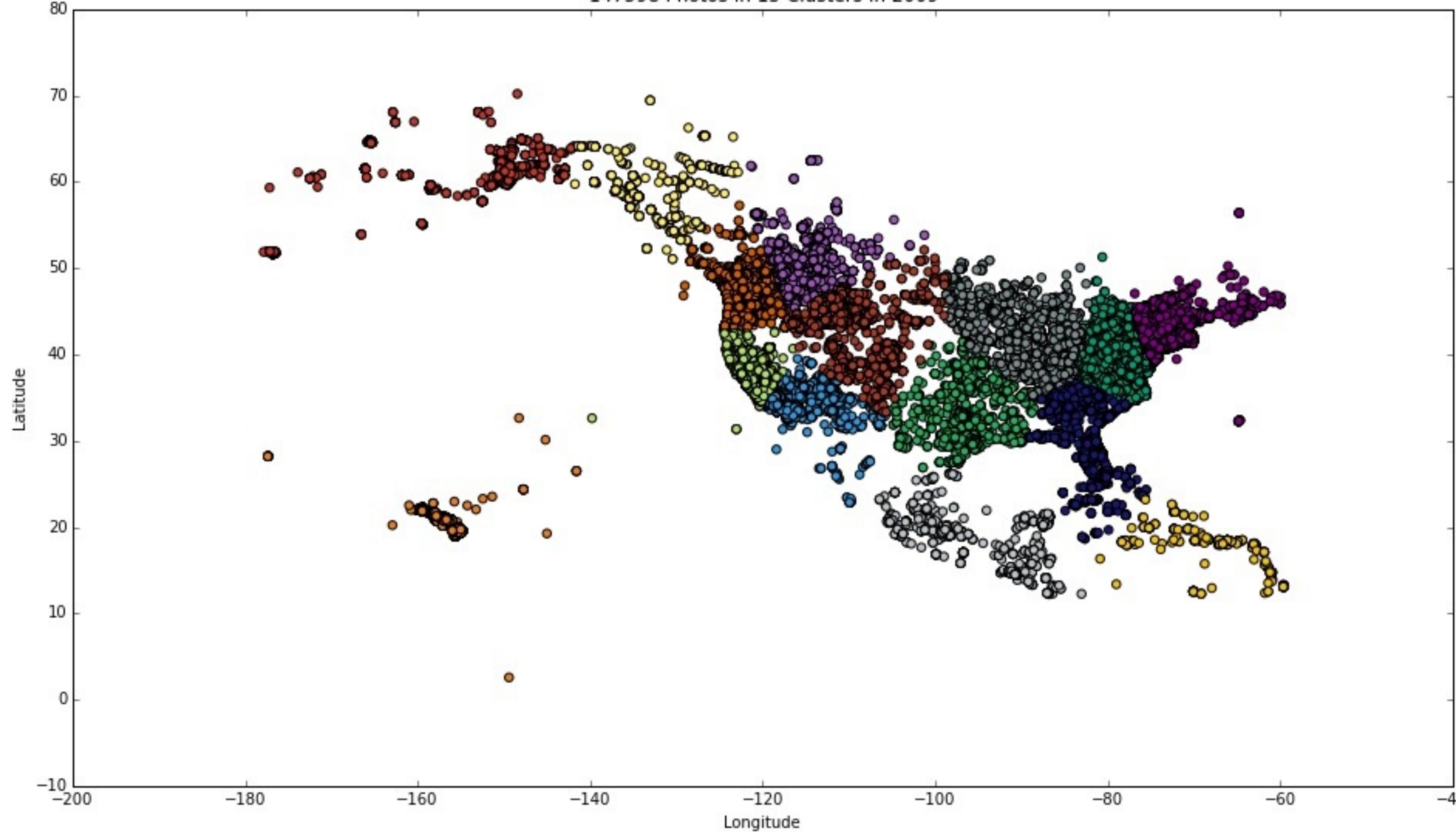
K-Means Clusters
Cluster Regions by Year

143535 Photos in 15 Clusters in 2008



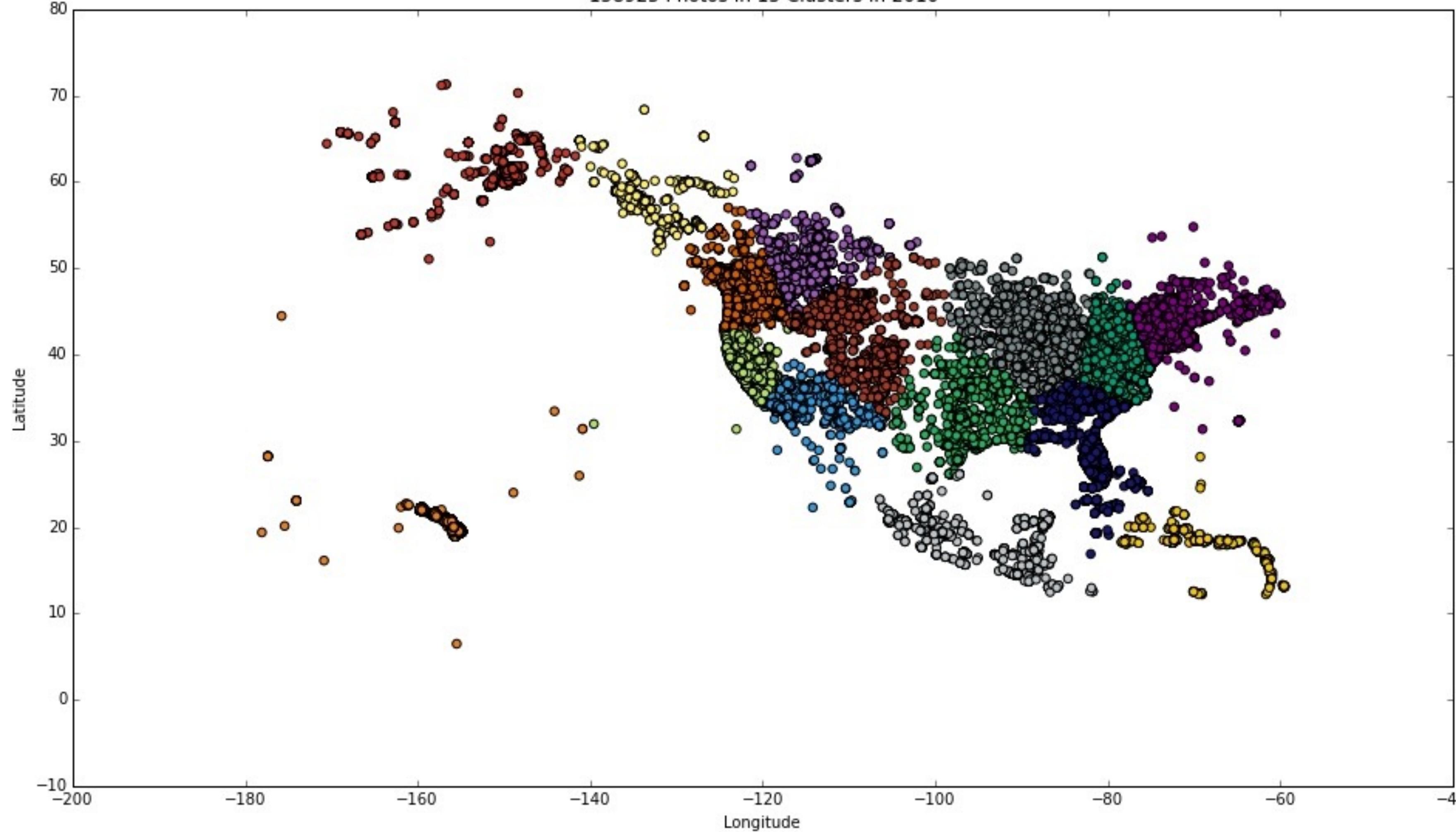
K-Means Clusters
Cluster Regions by Year

147598 Photos in 15 Clusters in 2009



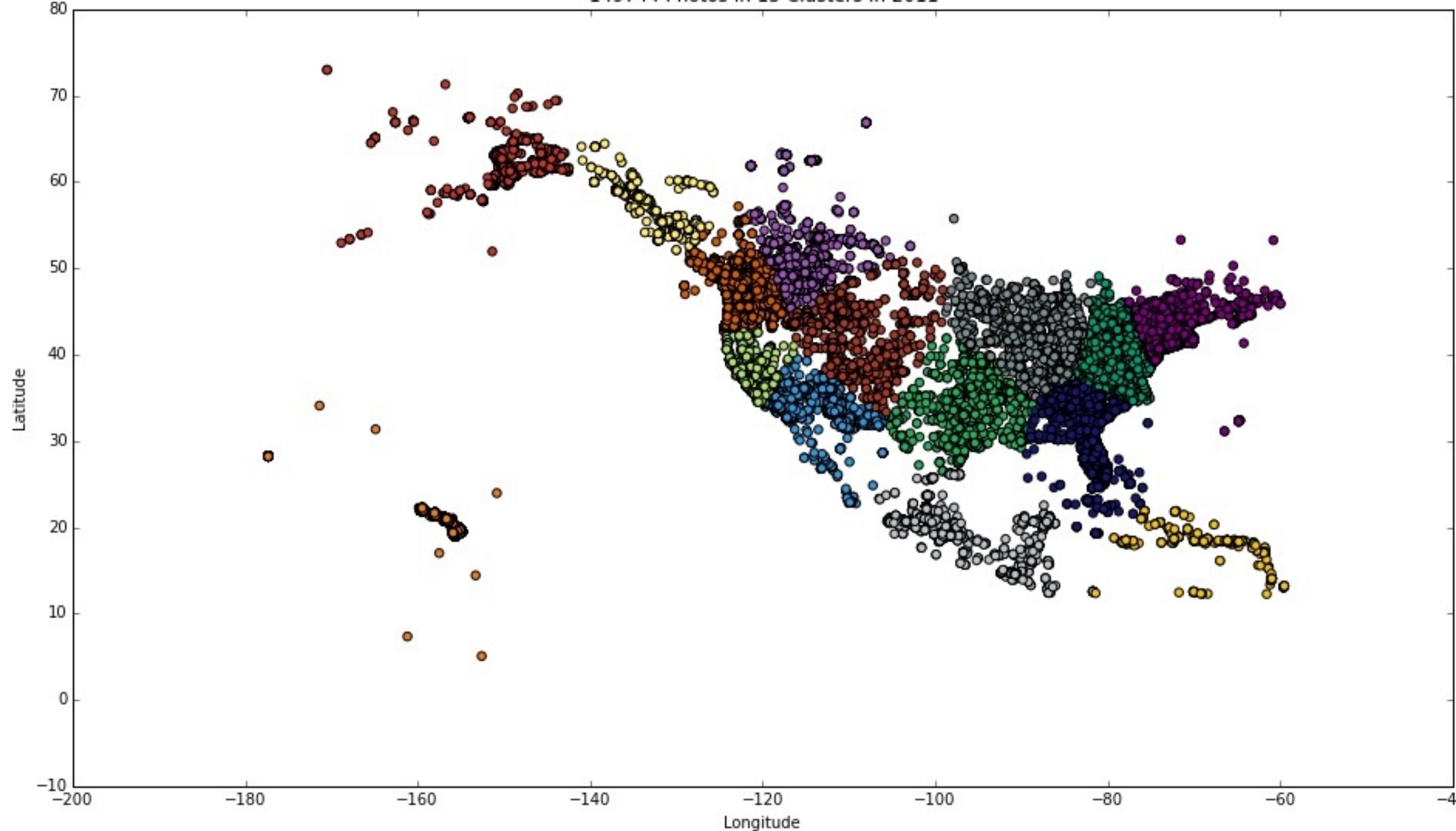
K-Means Clusters
Cluster Regions by Year

158925 Photos in 15 Clusters in 2010



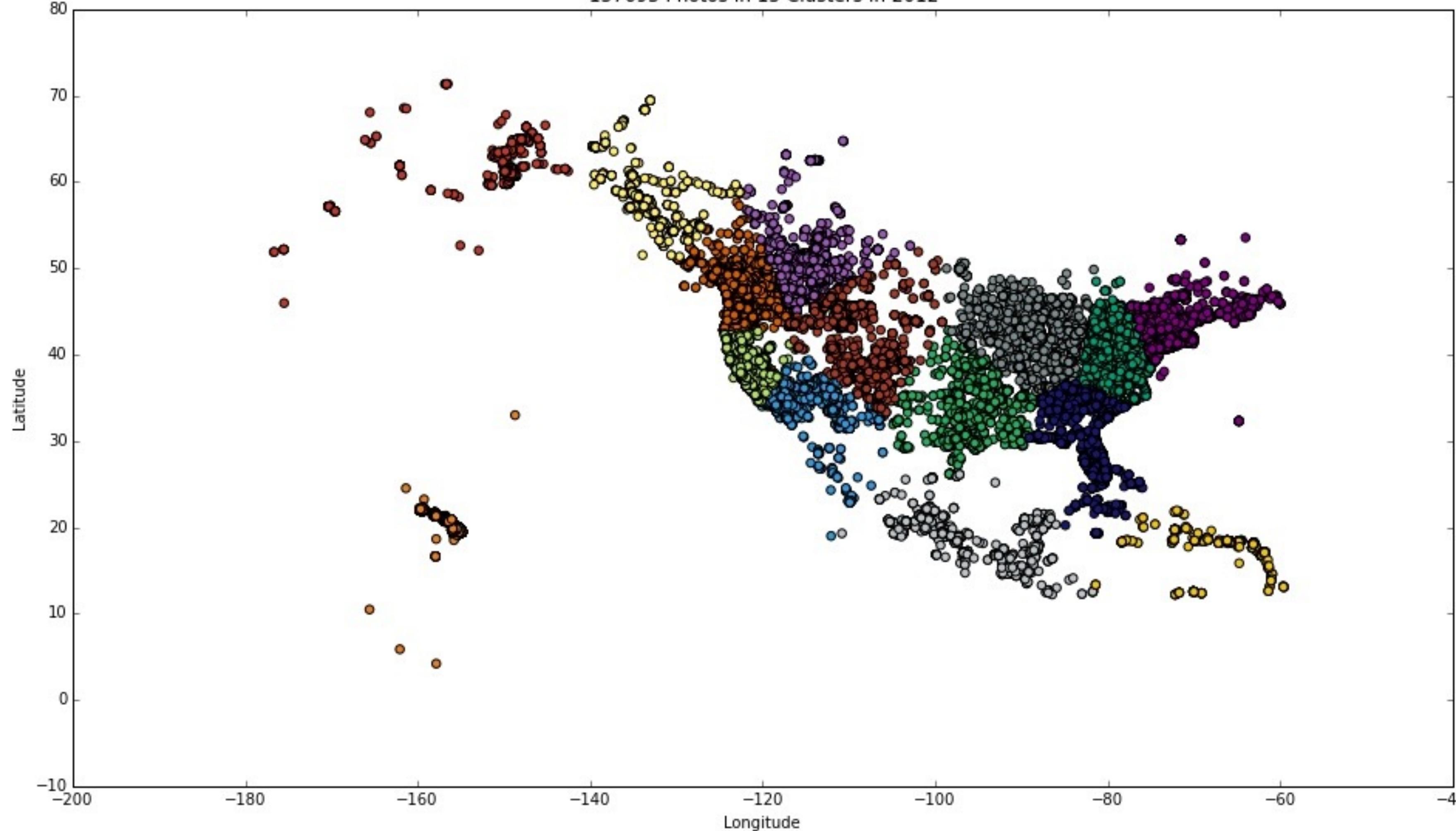
K-Means Clusters
Cluster Regions by Year

149744 Photos in 15 Clusters in 2011



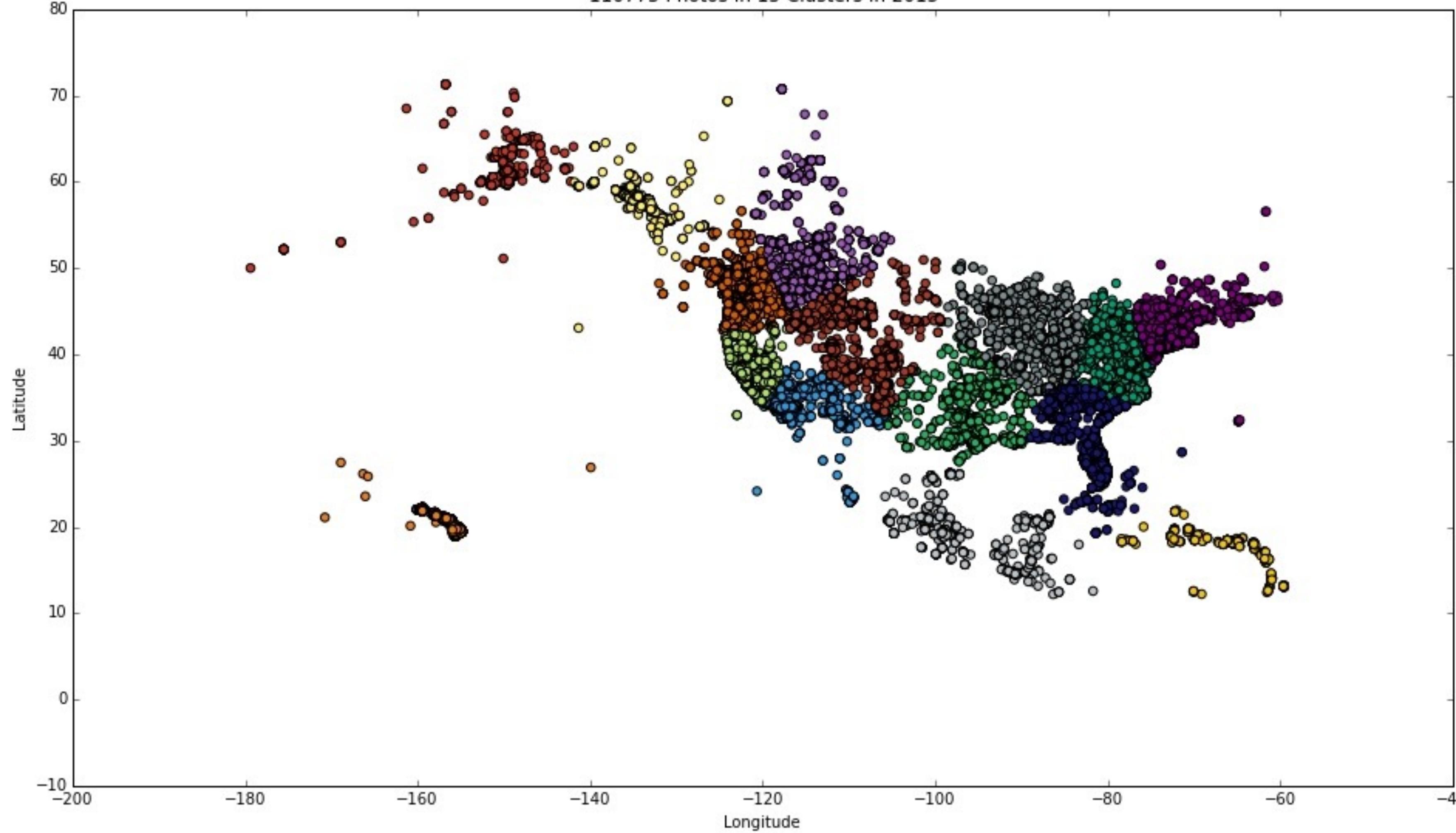
K-Means Clusters
Cluster Regions by Year

137693 Photos in 15 Clusters in 2012



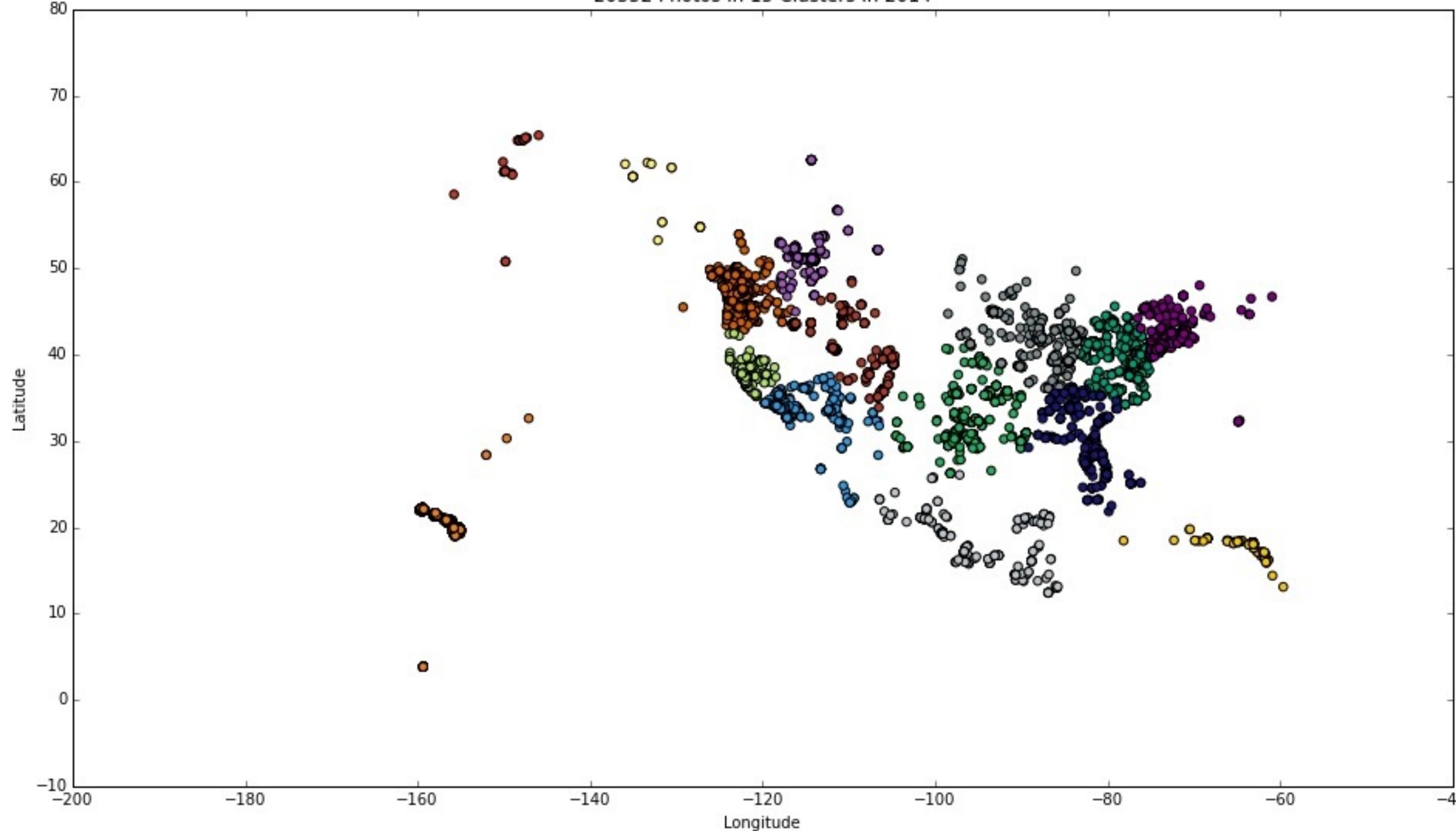
K-Means Clusters
Cluster Regions by Year

110773 Photos in 15 Clusters in 2013



K-Means Clusters
Cluster Regions by Year

20352 Photos in 15 Clusters in 2014



K-Means Clusters
Cluster Regions by Year

#	Sample	Train	Dataframe	x1	x2	x3	x4	x5	y
0	9	21	115	501	412	673			
1	21	115	501	412	673	2529			
2	115	501	412	673	2529	3029			
3	501	412	673	2529	3029	3183			
4	412	673	2529	3029	3183	4819			
5	673	2529	3029	3183	4819	4258			
6	2529	3029	3183	4819	4258	3270			
7	3029	3183	4819	4258	3270	3635			

Setting Up Model

Years: 2000 - 2013, since 2014 is incomplete

5-Year Slices

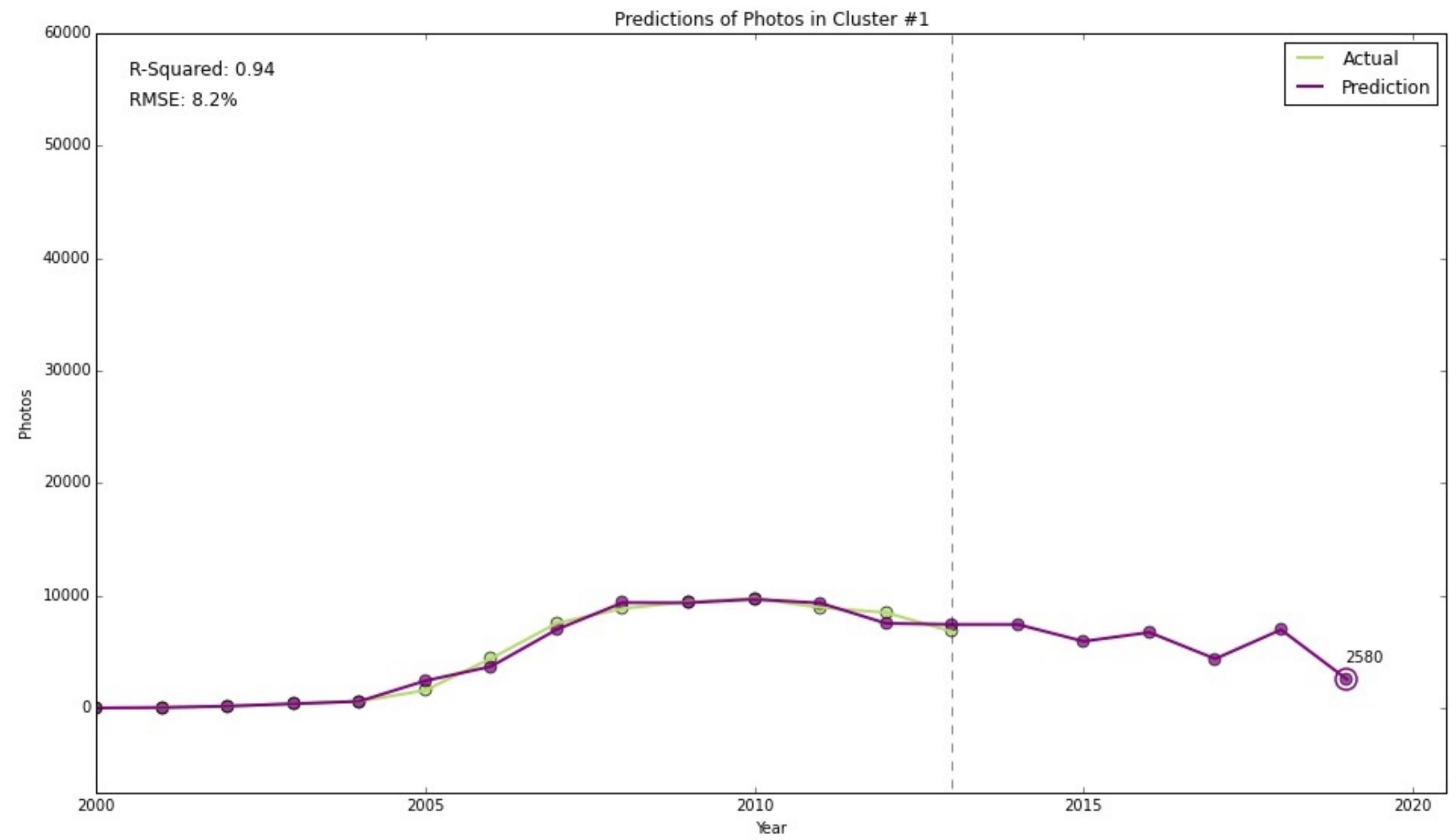
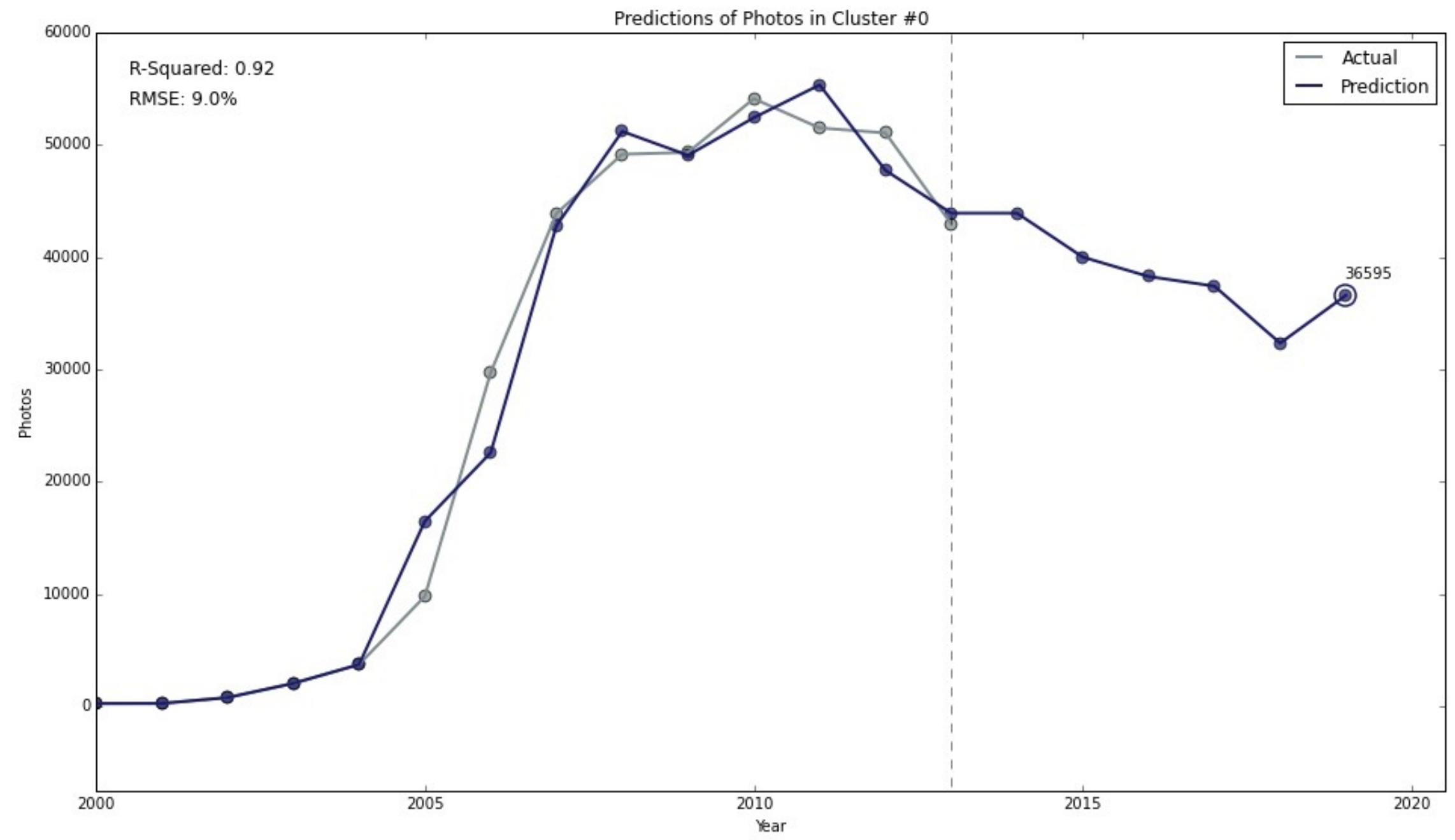
Time Series

Creating Dataframes

```
# Sample Testing Dataframe
```

	x1	x2	x3	x4	x5	y
0	9.000	21.000	115.000	501.000	412.000	1334.476
1	21.000	115.000	501.000	412.000	673.000	1917.193
2	115.000	501.000	412.000	673.000	2529.000	3190.911
3	501.000	412.000	673.000	2529.000	3029.000	3062.124
4	412.000	673.000	2529.000	3029.000	3183.000	4374.583
5	673.000	2529.000	3029.000	3183.000	4819.000	4601.086
6	2529.000	3029.000	3183.000	4819.000	4258.000	3222.234
7	3029.000	3183.000	4819.000	4258.000	3270.000	4050.724
8	3183.000	4819.000	4258.000	3270.000	3635.000	2856.667
9	4819.000	4258.000	3270.000	3635.000	2856.667	2856.667
10	4258.000	3270.000	3635.000	2856.667	2856.667	1935.043
11	3270.000	3635.000	2856.667	2856.667	1935.043	3570.485
12	3635.000	2856.667	2856.667	1935.043	3570.485	1288.946
13	2856.667	2856.667	1935.043	3570.485	1288.946	4346.693

Time Series
Creating Dataframes

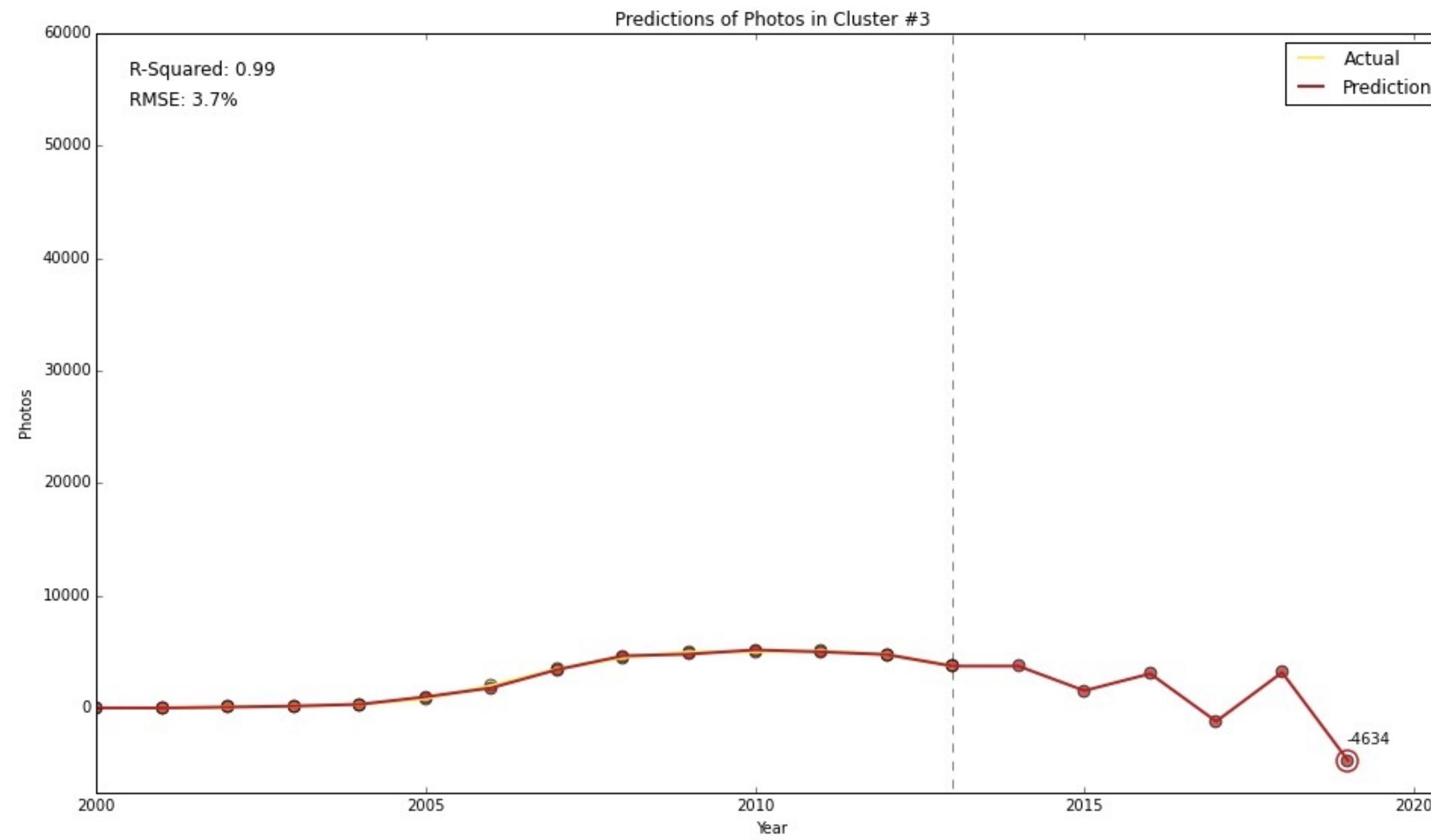
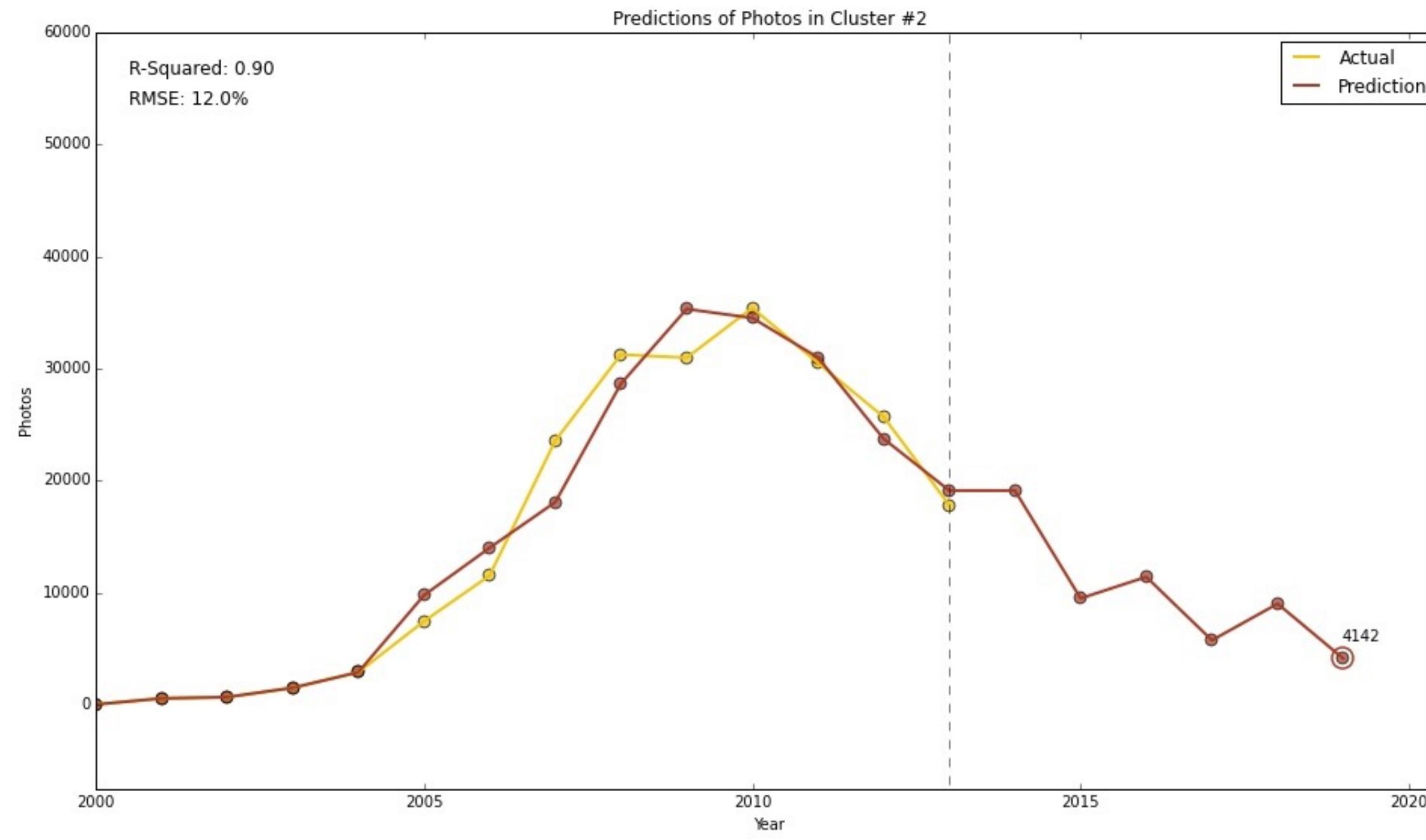


Accuracy Metrics

$R^2 = 86.2\%$

RMSE = 11.9%

Linear Regression
Predictions

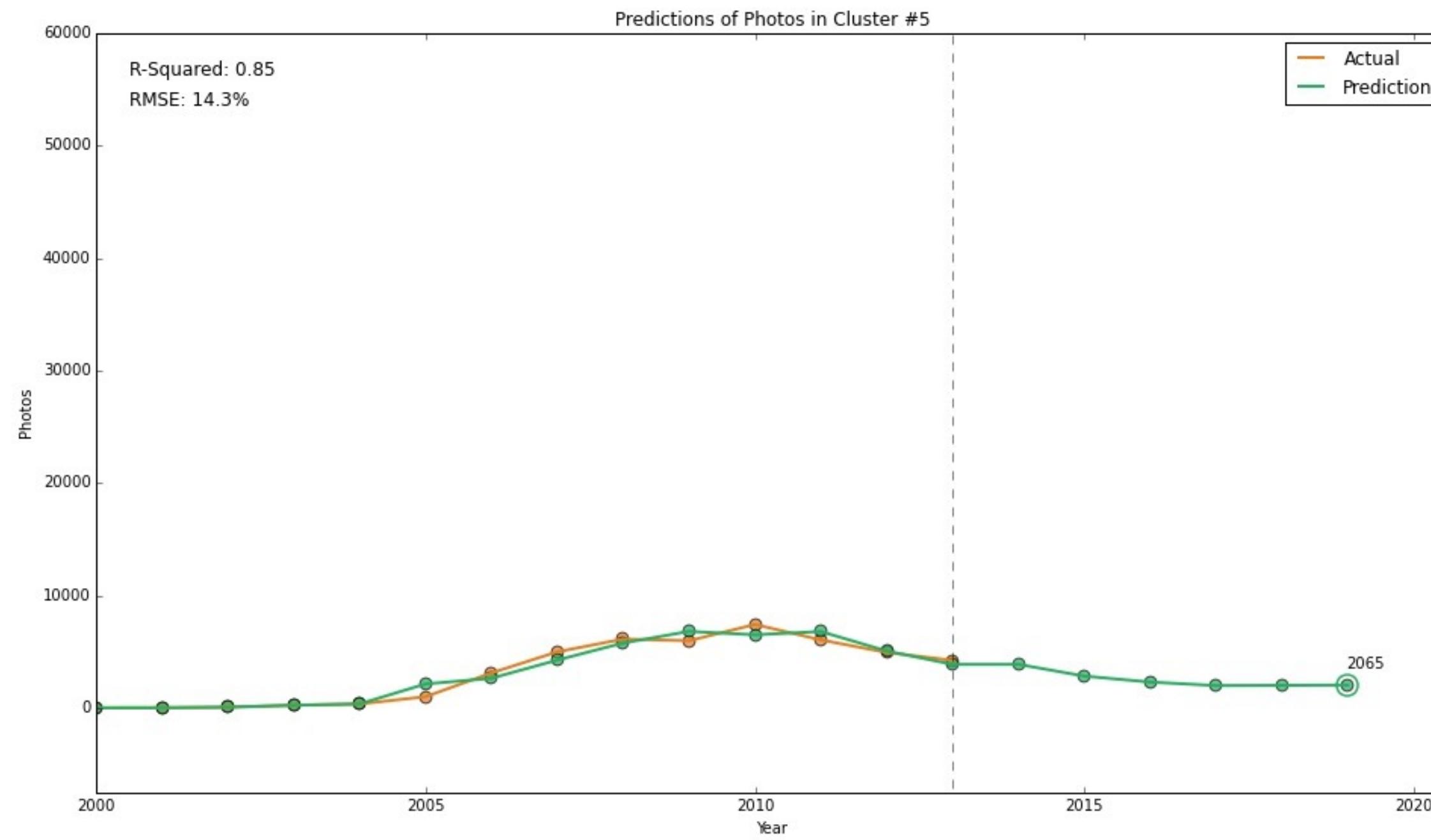
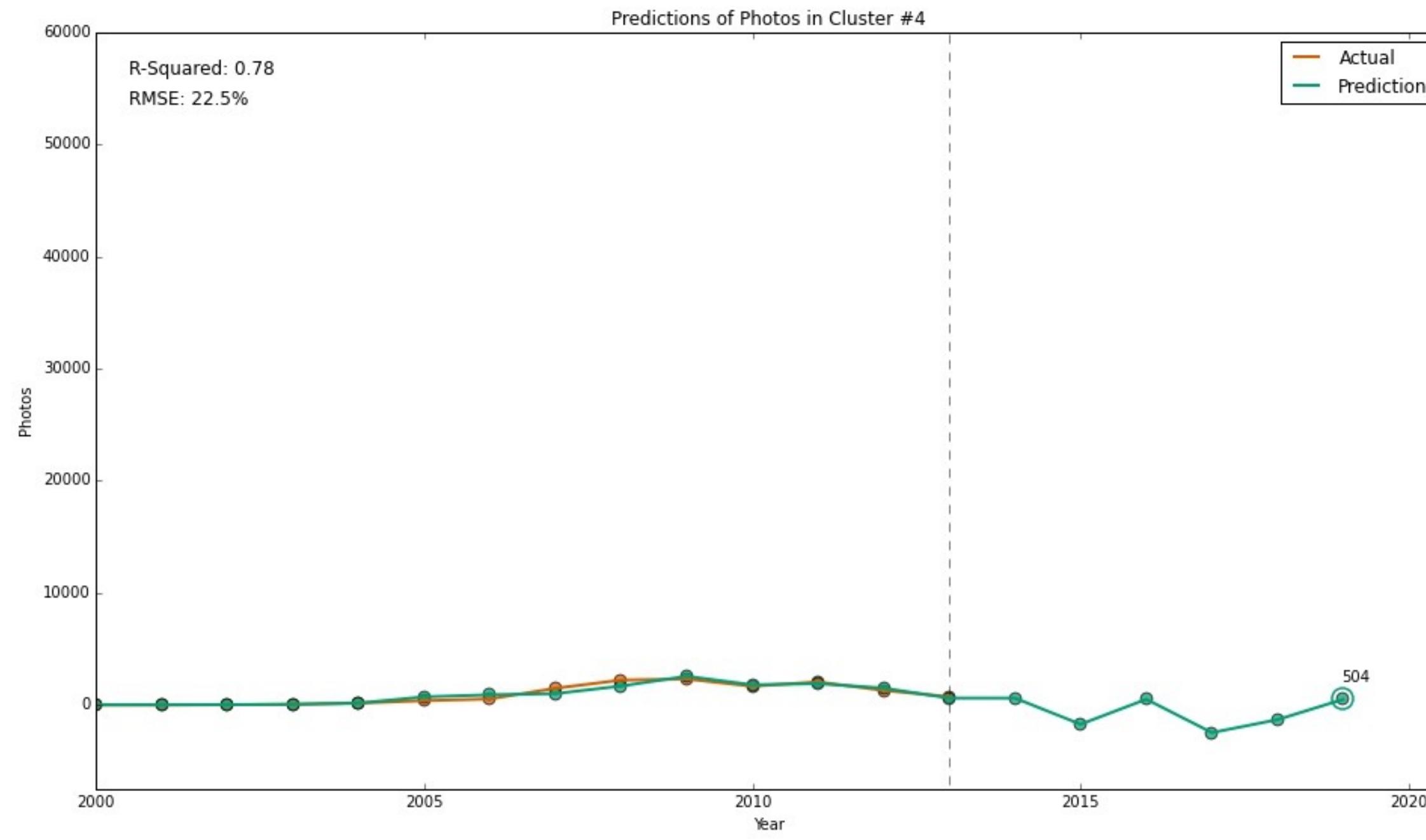


Accuracy Metrics

$R^2 = 86.2\%$

$RMSE = 11.9\%$

**Linear Regression
Predictions**

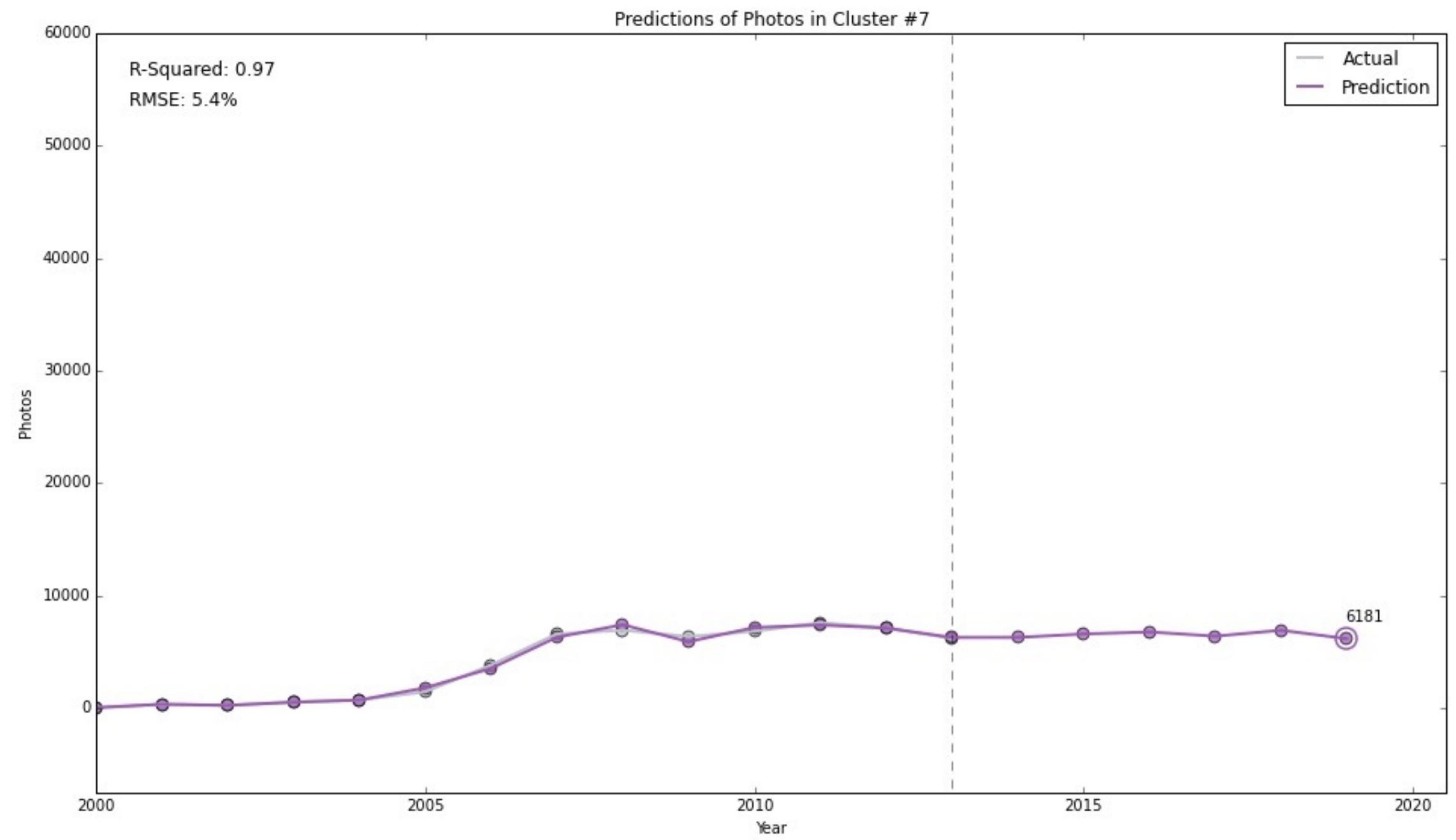
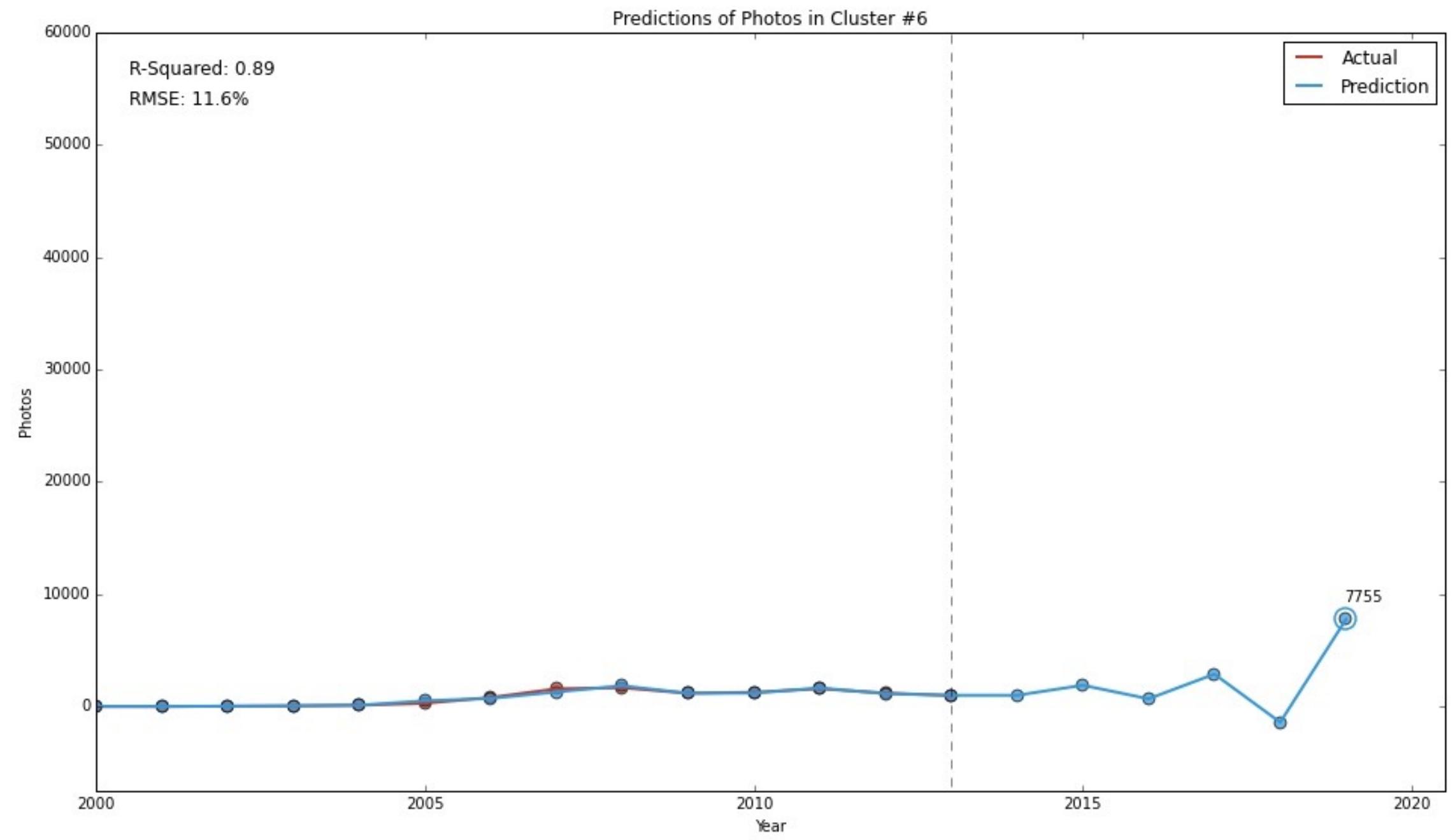


Accuracy Metrics

$R^2 = 86.2\%$

RMSE = 11.9%

**Linear Regression
Predictions**

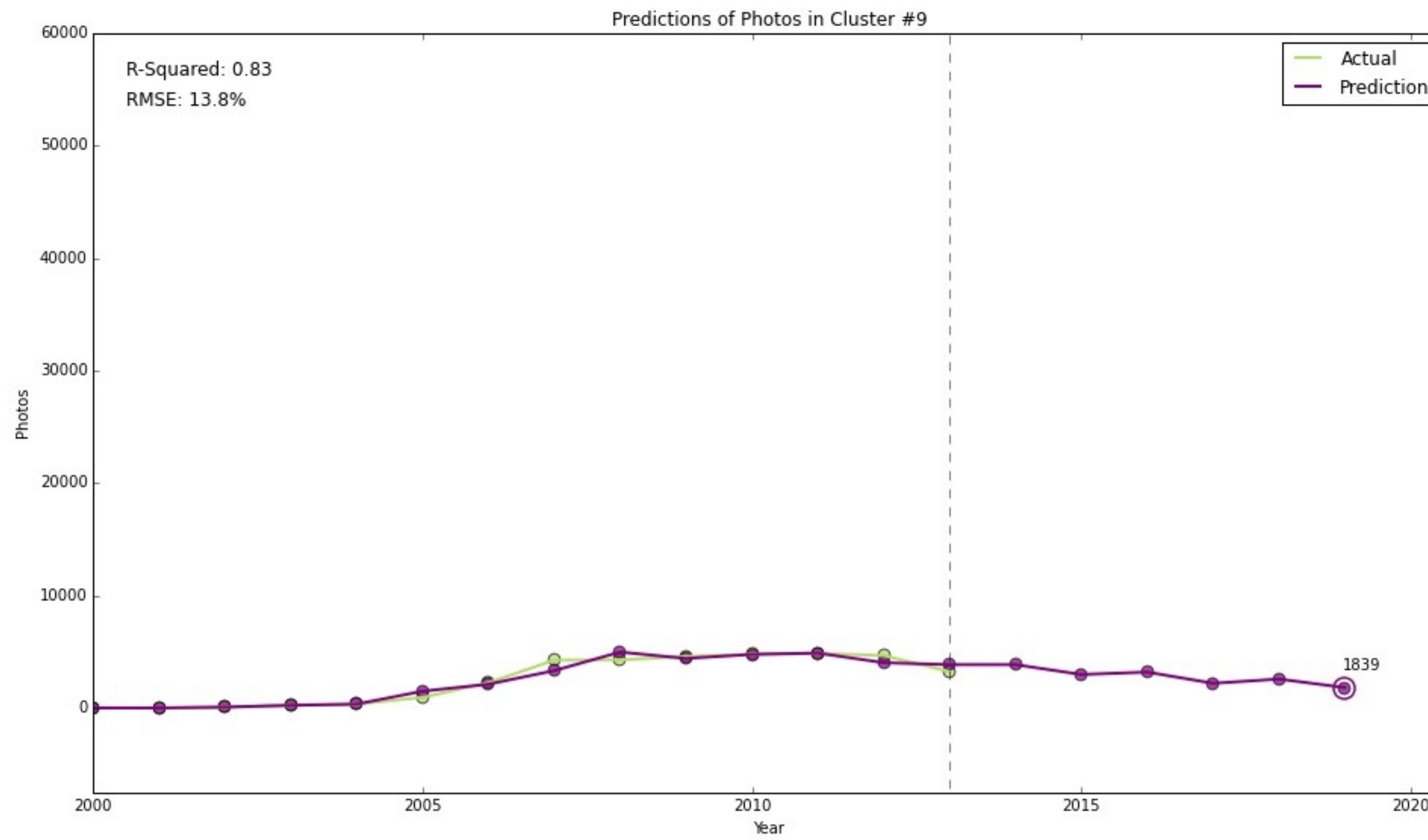
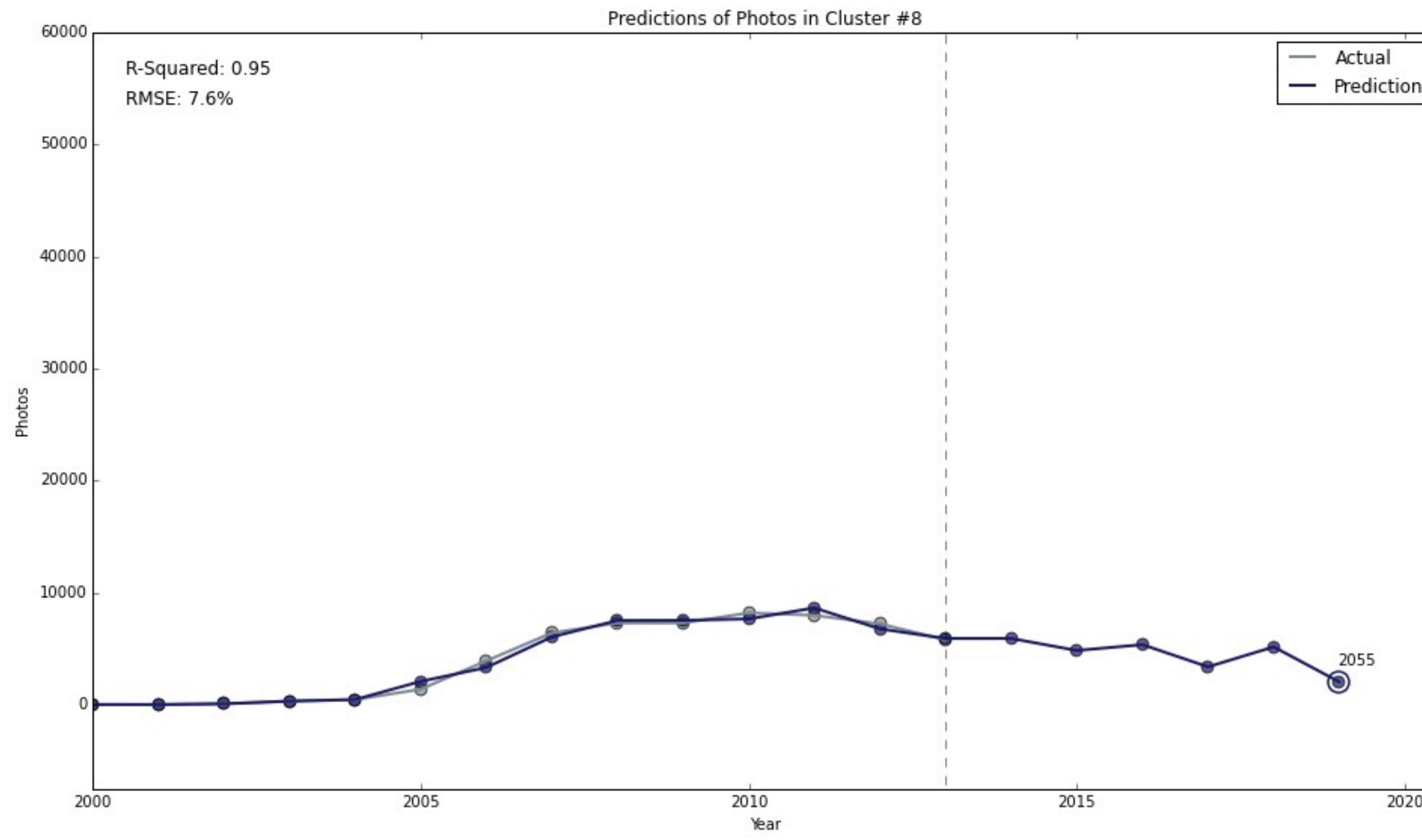


Accuracy Metrics

$R^2 = 86.2\%$

RMSE = 11.9%

Linear Regression
Predictions

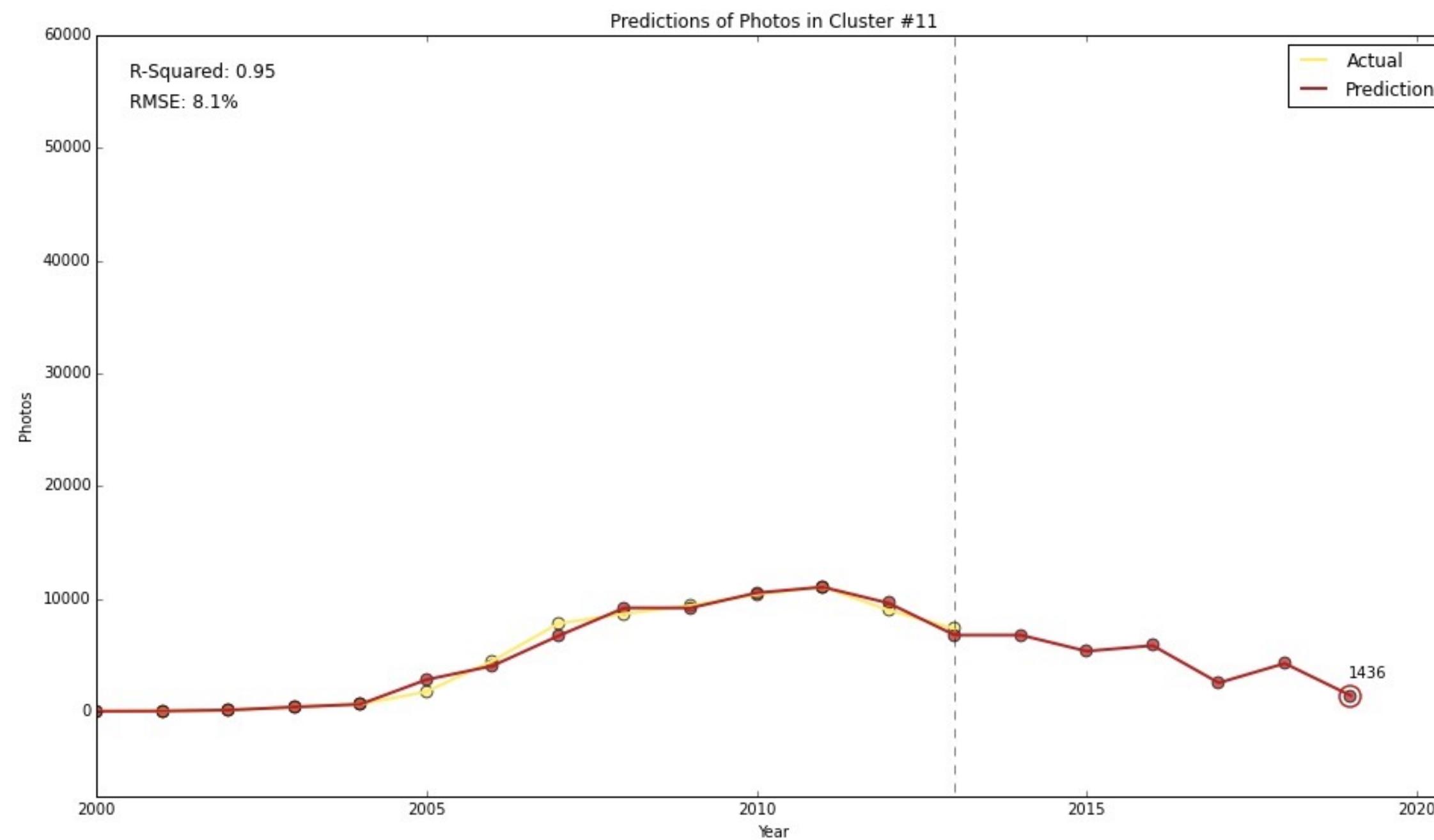
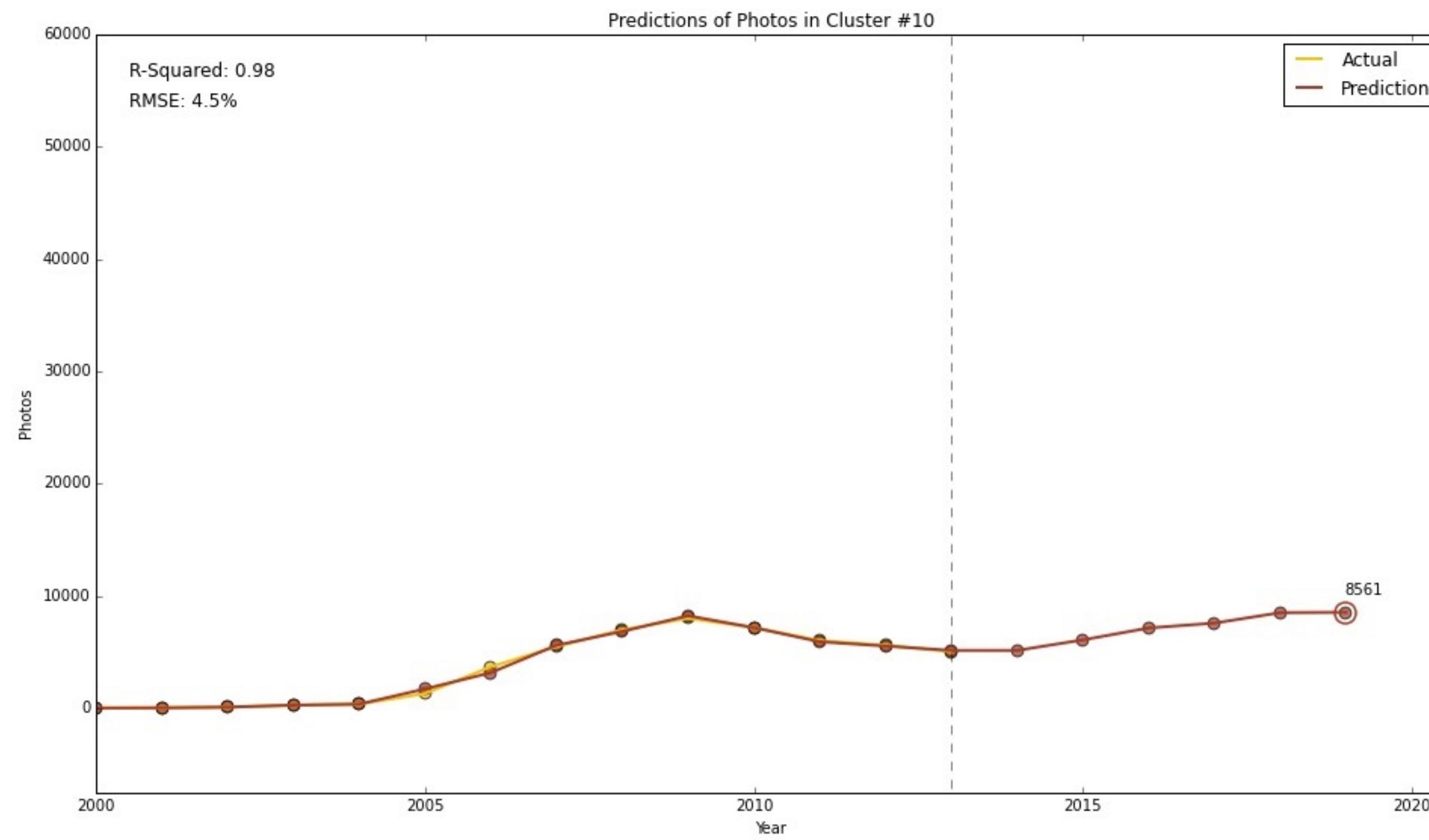


Accuracy Metrics

$R^2 = 86.2\%$

$RMSE = 11.9\%$

**Linear Regression
Predictions**

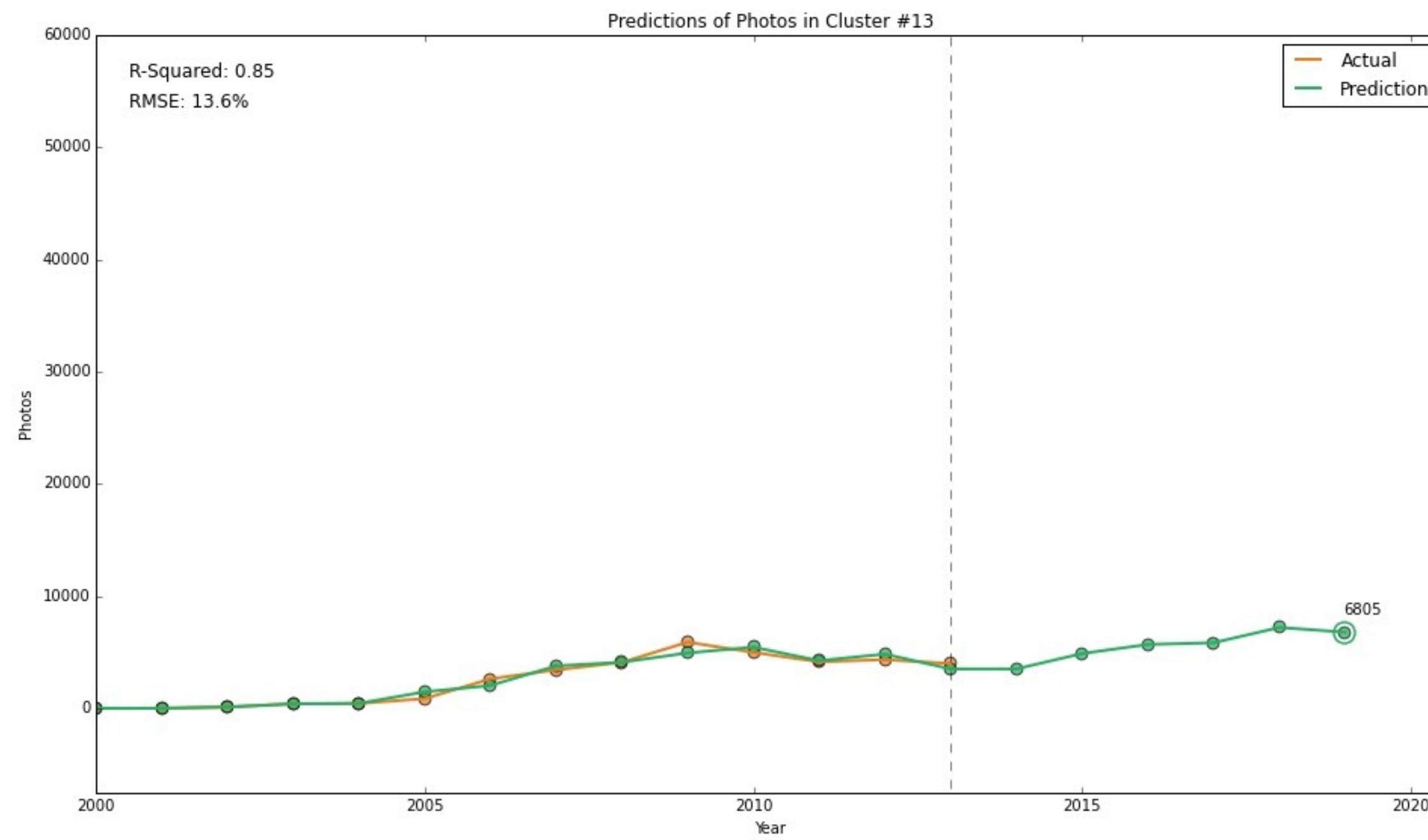
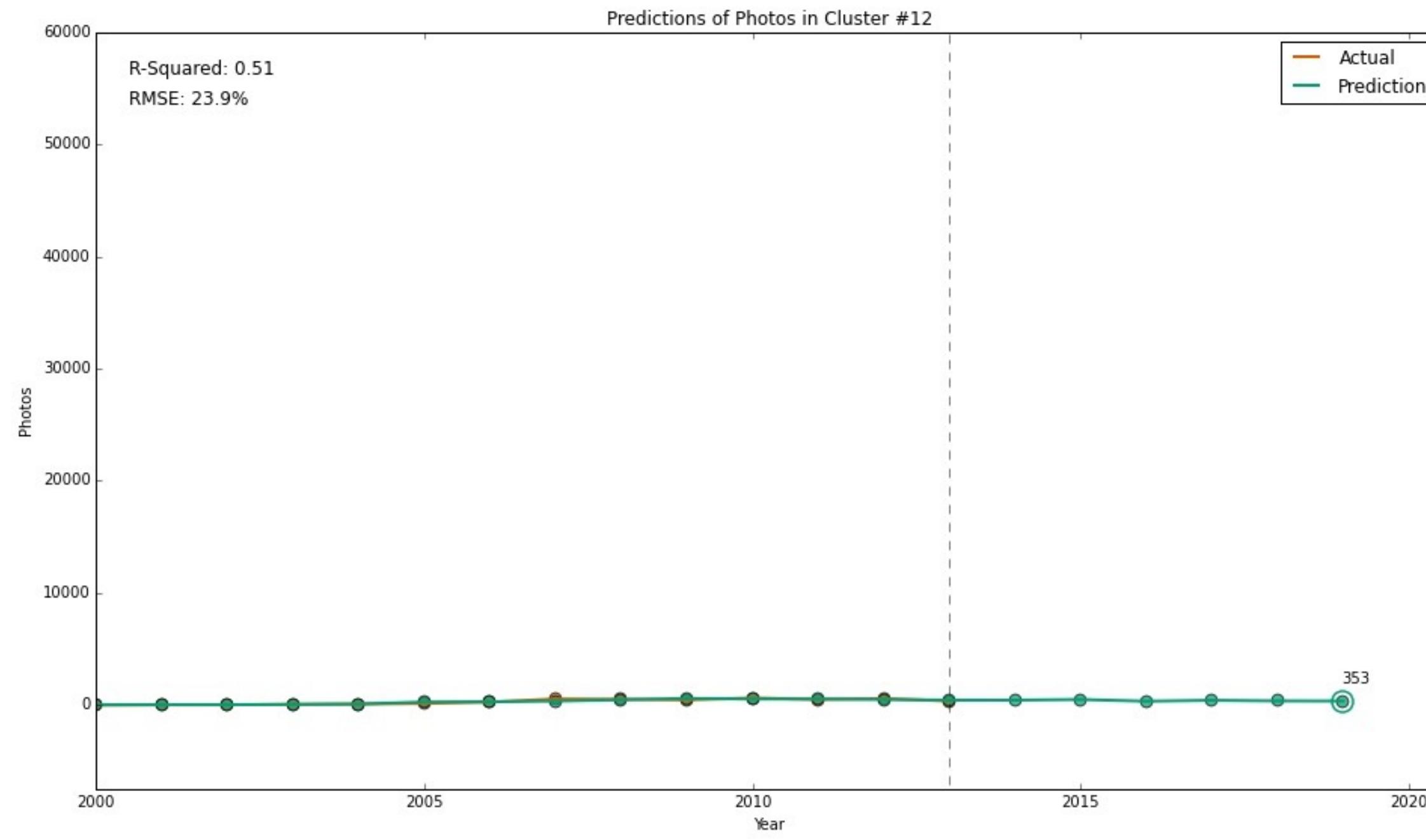


Accuracy Metrics

$R^2 = 86.2\%$

$RMSE = 11.9\%$

**Linear Regression
Predictions**

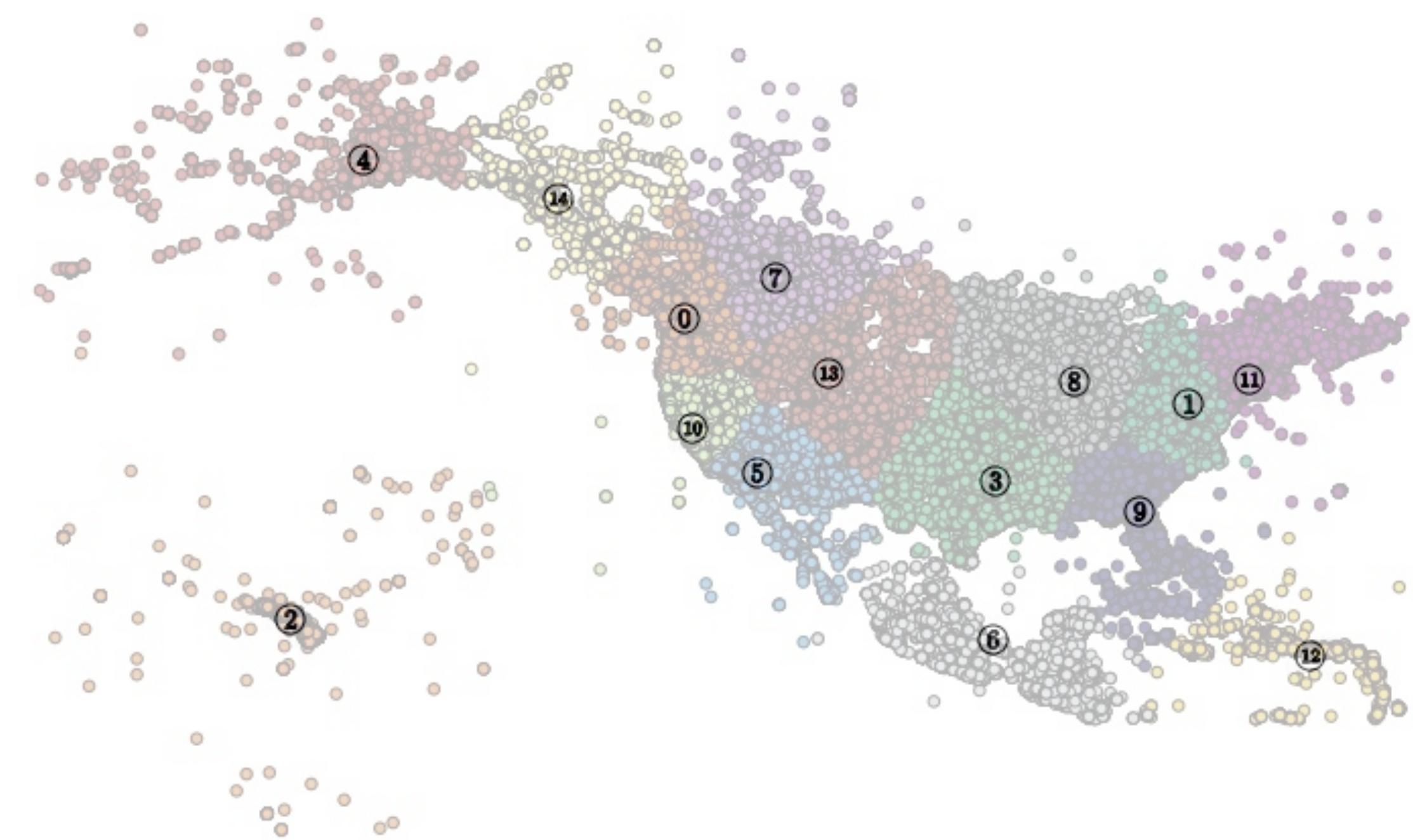
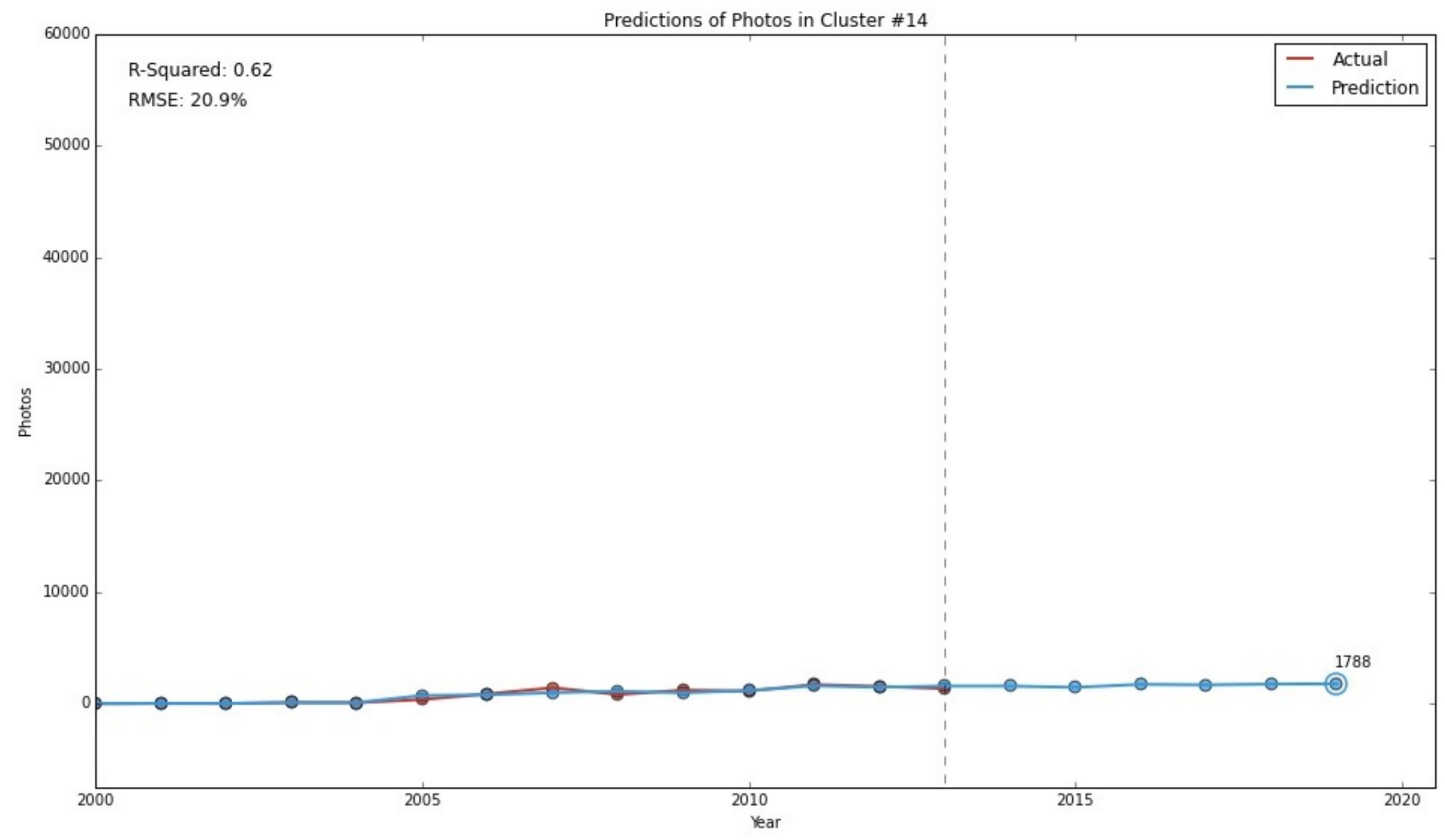


Accuracy Metrics

$R^2 = 86.2\%$

$RMSE = 11.9\%$

**Linear Regression
Predictions**

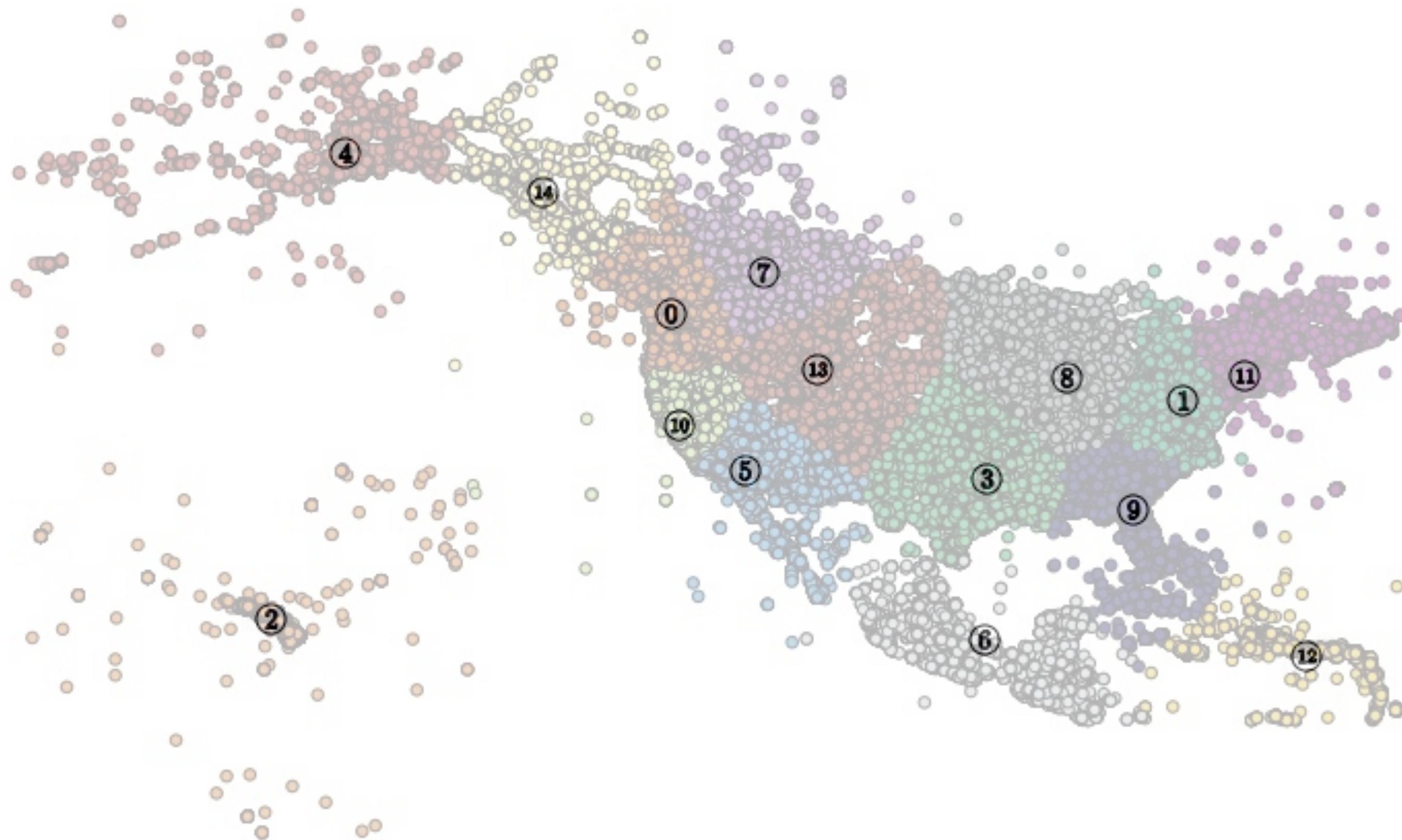


Accuracy Metrics

$R^2 = 86.2\%$

RMSE = 11.9%

**Linear Regression
Predictions**



Predictions for 2019

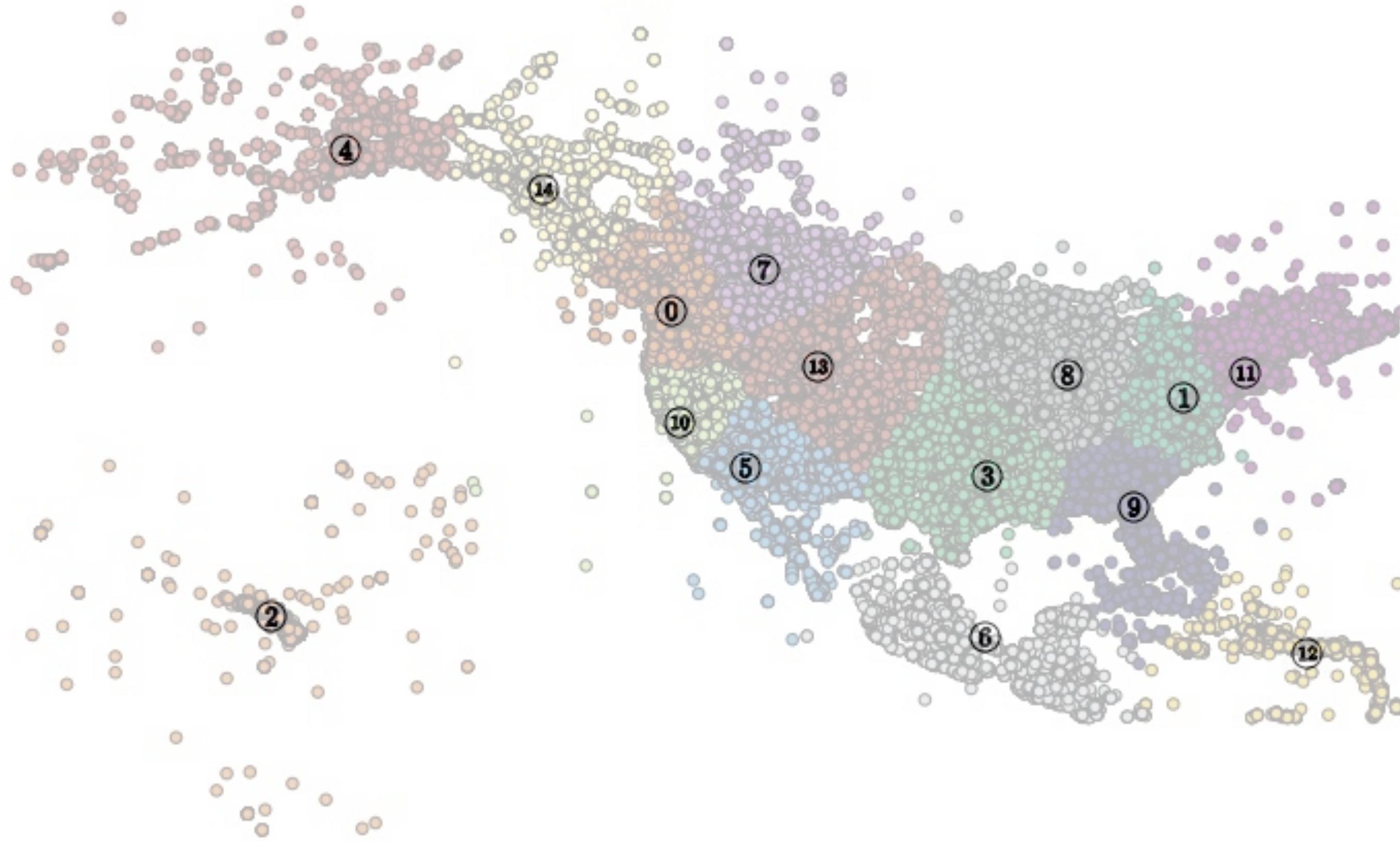
0: 36595	8: 2055
1: 2580	9: 1839
2: 4142	10: 8561
3: -4634	11: 1436
4: 504	12: 353
5: 2065	13: 6805
6: 7755	14: 1788
7: 6181	

Total: ~82659

*excluding negative values

Final Conclusions
Predictions

Popular Places in 2019



Most Popular:
Pacific Northwest

Least Popular:
Hawaii, Texas/
Midwest

Trending:
Central America/
Mexico

Final Conclusions
Predictions

- Huge dataset
- Processing power
- Time series slicing
- Accurate regression model

Problems

- Extrapolate to the World
- Other Models to Test Accuracy
 - K-Medians
 - SVM?
- Optimizing Clustering
- Use Tags to Isolate Photo Features
- Find Economic Status of Locations and Compare
- ... and more!

Next Steps